경영학석사학위논문

Do Insurance Firms Learn from Repeated Contracts?

- Evidence from the Korean Auto Insurance Market -

보험사의 정보 독점은 존재하는가?
–한국 자동차보험시장을 중심으로–

2016년 2월

서울대학교 대학원
경영학과 재무금융 전공
한 상 은

Do Insurance Firms Learn from Repeated Contracts?

- Evidence from the Korean Auto Insurance Market -

Sangeun Han

College of Business Administration

The Graduate School

Seoul National University

Keywords: asymmetric learning, informational monopoly, auto insurance, residual market

Student Number: 2014-20460

Contents

Tables and Figures

Abstract

With a large data set from Korean auto insurance market, this article examines the relationship between policyholders' risk and the age of policies. Claim risk decreases with policy age, especially due to the decrease in claim frequency as contracts are repeated. Findings are consistent with theory on asymmetric learning and support the existence of an informational monopoly of an incumbent insurer in the Korean auto insurance market. Investigation on policies in the residual market further implies the strategic behavior of insurance companies based on their informational advantage.

1. Introduction

Studies on informational asymmetry have continuously evolved ever since the work of Rotschild and Stiglitz (1976). A large part of research focused on the cross-sectional relationship between policyholders' risk and insurance coverage to examine the existence of adverse selection or moral hazard. However, information is not only a static factor and can be generated over time. Investigation on asymmetric learning catches these dynamic aspects of information and considers repeated contracting situations among two parties. The party who gains from asymmetric learning wins market power and subsequently generates profits.

Despite the importance of information asymmetry and asymmetric learning in the empirical aspect, theory was not sufficiently supported by empirical findings until the last decade. After the findings of Chiappori and Salanie (2000) that insurance data well suit the empirical setting for examining informational asymmetry, a series of studies investigated the existence of asymmetric information in various insurance market situations. In this article, I extend the literature one step further by identifying whether asymmetric learning exists in the Korean auto insurance market.

The fundamental idea of this article is whether insurance companies build an informational monopoly by learning about their policyholders and exploit their informational advantage to generate profits. As a means to control loss ratios, insurance companies make efforts to learn about their policyholders by strengthening their underwriting and actuarial functions. It becomes an issue if these activities become a property to the incumbent insurer. This is an important point because when incumbent insurers generate information that their rivals are unable to observe, they can take an advantage from their learned information. As contracts are renewed, insurers can adjust premiums differently that is, they can perform price discrimination in a way that the firm takes more profits from low-risk insureds.

The article is focused on the influence of the age of policies. First, whether the average risk of policyholders decreases with policy tenure will be one main point. When insurance firms gain an informational advantage they will apply a strategy to retain a larger fraction of low-risk insureds and a relatively smaller fraction of high-risk insureds. With asymmetric learning, informational monopoly grows with policy age and as a result, a negative relation between policy age and the average risk of policyholders is established. Second, a more direct method to identify whether incumbent insurance companies gain from informational advantage is to test the relationship between profits and policy age. Profits should be higher for repeated contracts.

The comprehensive data set used in this article makes it possible to account for individual-level information. This is an extension in the field of study since most of the insurance data use aggregated data sets which have limited intuitions. Furthermore, I investigated policy performance of policies in the residual market (shared market, involuntary market); high-risk policies which are mandatorily assigned to insurers. The comparison between the voluntary and involuntary market implies the strategic behavior of insurance companies since the main difference between the two policy groups is the insurers' ability to control risk by voluntarily deciding whether to write the contract or not.

The remainder of this article is structured as follows. Related literature and hypotheses development are explained in the next section. Section3 describes details of the comprehensive data set and the unconditional relationship between policy age and risk. Section 4 and 5 present the framework of study and examine the main results on the relationship between policy age and policyholders' risk and insurers' profitability. Section 6 concludes.

2. Related Literature and Hypotheses Development

Related Literature

Several studies investigated the existence of asymmetric learning in the insurance market. Mainly based on the adverse selection model of Rothshild and Stiglitz (1976), Kunreuther and Pauly (1985) concluded that insurers exploit their information monopoly and therefore, prices and profits increase over time in a setting where incumbent insurers are able to keep their information private. Multiple studies extending previous models include D'Arcy and Doherty (1990) and Dionne and Doherty (1994). Both found evidence of insurers' behavior to systematically overcharge new polices and use commitment in order to low-risk policyholders. These studies used an aggregate data set which makes it difficult to identify the effect on the individual level. The large set of individual level data used in this article makes it possible to overcome this shortcoming and investigate the auto insurance market more carefully. Chiappori and Salanie (2000) show a link between theory and empirical evidence of asymmetric information and provide an empirical application on the French automobile insurance market where no evidence for asymmetric information is found.

This article is closest to Cohen (2012) and Kofman and Nini (2013). Cohen (2012) investigates the temporal pattern of profitability in a panel of Israeli insurers and concludes that there is evidence of asymmetric learning in the Israeli automobile insurance market. This conclusion is drawn from the fact that the ratio of premium to losses is larger for policyholders with repeated contracts. On the other hand, Kofman and Nini (2013) used a large data set of Australian automobile insurance market and found no evidence for asymmetric learning. Both studies controlled for observable characteristics such as age, gender, or vehicle age. There are various possible causes for the different results between the two studies but I conjecture that the existence of bonus-malus coefficient plays an essential role in whether insurers possess an informational monopoly. A bonus-malus coefficient reflects the claim history of individual policyholders and is shared by insurers. By sharing bonus-malus records, information about policyholders is no longer held private by the incumbent insurer and subsequently, claim history becomes an observable characteristic that is easily obtainable.

The Israeli insurance market does not regulate insurers to share information through bonus-malus systems. That is, incumbent insurers who are able to observe the claim history of a policyholder as the contract is repeated, has a high chance to hold information monopoly. However, in Australia, a "no claims discount" (NCD) rating factor is shared among insurers which reflects the frequency or timing of claims made. Subsequently, incumbent insurers do only have an extremely restricted informational advantage compared to competitors.

In most countries that apply a bonus-malus scheme, the bonus-malus factor tracks the occurrence of at-fault claims. This is also the case in the Australian insurance market.

However, the Korean insurance market applies a unique way to build the bonus-malus factor. Unlike in other countries, the factor in Korea is based on accident severity rather than accident frequency. This may lead to a contamination in the bonus-malus coefficient since a series of small accidents may be given the same coefficient as an occurrence of a single accident with high severity. This implies that in the Korean insurance market, incumbent insurers who observe the exact claim history of repeated policyholders may generate informational monopoly. On the other hand, regulations conducted by the Korea Insurance Development Institute probably removes the possibility of informational monopoly since all individual level data is collected and shared. Even details on accidents, which are possible sources of informational monopoly, are known by all insurance firms after a certain period of time. Thus, I will examine empirical evidence to clarify whether information monopoly exists in the Korean auto insurance market.

Hypotheses Development

Kunreuther and Pauly(1985) and Kofman and Nini(2006) showed that in a situation where contracts are repeated, the incumbent insurer develops certain pricing power and exploits this to retain a higher percentage of low-risk policyholders. Kofman and Nini (2006) identified a positive relationship between claim occurrence and lapsation. Higher lapse rate for risky drivers seems to be consistent with asymmetric learning.

However, there are various possible explanations for such cross-sectional relationship between accident frequency and lapse rate other than asymmetric learning. Thus to confirm the existence of informational monopoly and pursuing consistency with former literature on policy lock-in, I will focus on the following two hypotheses. This hypotheses development is also consistent with Kofman and Nini (2013).

H1: Conditional on publicly observable information, insurance risk and the age of the policy are negatively related

To carefully observe the existence of informational monopoly, the control for observable characteristics is essential. Even if a certain character is closely correlated with both policy age and accident risk, when all insurance firms can observe it, it cannot be regarded as a source of monopoly. Thus, in this article, I control for various observable characteristics contained in the data set. Insurance risk is defined by both accident frequency and accident severity. I use the number of claims as a proxy for accident frequency and the claim amount for accident severity. Severity is measured conditionally; that is, claim amount given an accident is examined.

Based on theory, insurers will gain a higher profit from repeated contracts. Therefore, profitability and policy tenure have a positive relation. Further, the size of profits indicates the strength of informational monopoly of the incumbent insurer compared to competitors.

H2: Conditional on publicly observable information, insurance *profitability and the age of the policy are positively related.*

In short, the null hypothesis throughout the article is that neither average claim risk nor average profitability is related to the age of policy.

3. Data and Summary Statistics

The data set was obtained from one of the largest insurance companies in Korea. The individual level insurance panel data contains auto insurance claim records, coverage choice, premiums, and rating factors. Rating factors include variables such as policyholder's age, gender, the age and type of the car, its capacity, and region of registration. The claim records can be used as a proxy for accident information, including the number of claims and the amount of claim payment. I assume that the number of claims and the amount paid proxy accident frequency and severity respectively. The sample period includes data from the year 2009 to 2012 and the whole data set has 434,262 samples. Observations with less than one year of exposure were excluded in the frequency regression and policies that are co-insured are also left out for separate inspection. This leaves 410,287 observations. The dataset for the severity regression contains only 198,407 samples. Since I focus on the severity of a given accident, policies with partial exposure are also included but the sample set is smaller as only policies with claims are considered.

Table1 presents the definitions of all variables included in the dataset. Independent variables are classified into three categories describing policyholders, vehicles, or policies. Following Kofman and Nini (2013), I assume that all information available in the dataset is publicly available. Since the information below is collected at the start of the contract, this seems to be a reasonable assumption. The fact that certain information is publicly available means that it cannot be a source of information monopoly to the incumbent insurer against other competing insurance companies. Throughout the article, I will examine the impact of policy age on accident risk while controlling the publicly available variables. Especially noteworthy is the BMS variable, which reflects the claim history of a policyholder but is less reliable compared bonus-malus indicators in other countries for reasons described beforehand.

Unconditional Claim Risk and Profitability by Policy Age

Before starting the examination of policy age and accident risk, Figure1 shows the unconditional relationship among claim risk and profitability by policy age. As shown in the figure, the average loss decreases from about 260,000 Korean won to 92,000 Korean won monotonically as policy tenure gets older. Further, claim loss also shows a downward trend by policy age, ranging from 261,000 Korean won to 167,000 Korean won. It is observable that the slope is steeper for claim amounts than premium, which indicates the possibility of insurers generating profits from policies that claim less. The increasing profitability measure of (1-Loss Ratio) supports the possibility. Overall, unconditional averages imply that the average risk falls with policy tenure. Figure1 presents both the average premium and claim costs. Both trend down with policy age, and (1-Loss ratio) grows subsequently.

As mentioned in the "Hypotheses Development" part, to attribute the relationship presented in Figure1 to asymmetric learning, the relationship should remain after controlling for observable characteristics. That is, the patterns should not be explainable by observable factors. In the next section, I will examine the relationship in detail.

4. Empirical Setup

As a start for the empirical examination, I define the random variable $\widetilde{N_i}$ as the number of claims on policyholder *i*, and the random variable $\widetilde{S_i}$ as the severity of accident conditioned on occurrence. For accident frequency, I assume a Poisson distribution and for severity a lognormal distribution. Following Kofman and Nini (2013), both distributions account for unobserved heterogeneity.

The assumption for the claim frequency is that $\Pr(N) = \frac{e^{-\theta v} v\theta^N}{N!}$, where Θ is the expected value of the number of claims and v is a random variable which captures the unobserved heterogeneity in the model. In order to examine the annual number of claims, only samples with exposure of at least one year are used. The Poisson parameter varies by observable characteristics. In short, $\log(\Theta i)=X_i\beta_N+v$ is assumed where the random variable v is modeled as a gamma random variable with mean 1 and variance σ_v^2. As a result, the number of claims follows a negative binomial model as described below.

$$\Pr(N) = \frac{\Gamma(\frac{1}{\sigma_v^2}+N)}{\Gamma(\frac{1}{\sigma_v^2})\Gamma(1+N)}\left(\frac{\frac{1}{\sigma_v^2}}{\frac{1}{\sigma_v^2}+\theta}\right)^{\frac{1}{\sigma_v^2}}\left(\frac{\theta}{\frac{1}{\sigma_v^2}+\theta}\right)^{N} \quad \text{..................} (1)$$

The estimates and σ_v^2 are estimated by maximum likelihood estimation. From the equation, σ_v^2 indicates the size of unobserved heterogeneity. In the case where σ_v^2=0, the frequency distribution becomes a standard Poisson distribution.

For the severity model, I assume a lognormal distribution and account for both claim and person.

$$\ln(S_{ij})=X`_i\beta_s + \varepsilon_{ij} + \mu_i \quad \text{..................................} (2)$$

The random variable μ presents the unobserved heterogeneity. ε_{ij}, a random variable with a mean of zero and variance σ_ε^2 presents the standard regression residual.. σ_μ^2 indicates the heterogeneity on policyholder level. Due to this setting, this random variable directly captures the individual random effect in the claim severity regression. The parameters can be estimated by maximum likelihood.

Throughout the article, I use three different specifications and include various control variables. All the control variables used in the regressions are presented in Table2 and all have a significant relation with claim risk which is proven before starting the examination.[1] The three specifications differ in flexibility. Specification (A) includes a full set of control variables and the policy age dummies from one year to four years or longer. Specification (B) includes only the dummy variable for new policies and (C) does not contain any variable controlling for policy age.

5. Empirical Results

Policy Age and Claim Risk

Claim Frequency Results

Figure2 plots the actual claim frequency risk indicated by the thin black line (A), with the mean predicted values of both specifications (B) and (C). Actual claim frequency indicates the mean predicted value from the negative binomial model. Model (B) includes a dummy for new policies and (C) includes no controls for policy age. It is observable that the decrease in claim frequency risk by policy tenure cannot be entirely explained by observable variables. Observable characteristics that vary by age cannot sufficiently explain the decline in the actual frequency of claims as the policy gets older. The gap between actual claim frequency, specification (B), and (C) confirm the results.

Table3 reports the estimated coefficients from the three models respectively and presents the results in statistics. After controlling for observable characteristics, claim frequency risk decreases monotonically as policy age increases. A monotonic decrease in coefficients of policy age

[1] Results are not reported but available on request

variables is observed. Besides, each specification includes σ_μ^2 which measures the unobserved heterogeneity and all control variables listed in table2. It is noticeable that unobserved heterogeneity is significantly larger than zero and the size of estimate decreases by specifications. Strictly controlling for policy age increases the fit of the model. The table also reports measures of goodness of fit for each model. Both AICC and BIC become smaller with the inclusion of policy age controls. That is, adding more controls for policy age improves the fit of the model. Given the fact that two times the difference in log-likelihood values follows a chi-square distribution with degrees of freedom equal to the difference in numbers of parameters, comparing the goodness of fit by log likelihood ratio between model (A) and (B) also confirms increased goodness of fit of model (A). The likelihood ratio statistic is 68.42 where a five percent critical value of the distribution for 4 degrees of freedom is 9.49. In addition, comparing model (B) and (C), the dummy variable for new policies (policy age 0) is also statistically significant. The positive coefficient indicates a higher risk for new policies. Moreover, the likelihood statistic is 140.56, where the 0.5 percent critical value is 3.84. In short, from figure2 and table3 one can conclude that there exists informational monopoly in claim frequency in Korean auto insurance market and that new policies are riskier than others.

Claim Severity Results

The sequence applied in the claim frequency examination is repeated for the claim severity analysis. However, the sample size shrinks to 198,407. The sample set is much smaller than in the frequency analysis since claim severity is only examined when a loss occurs. In the claim severity regression, I account for a random effect for individual policyholders.

Figure3 shows that claim severity does not systematically decreases by policy age. It seems hard to predict claim amount with policy age which implies no informational monopoly gathered in the aspect of claim severity as contracts are repeatedly written. Further, observable characteristics cannot explain claim severity by policy age. In figure3, means of the predicted values calculated by each specification are presented. The actual claim severity which is indicated by the black thin line is not sufficiently explained by specification (B) and (C). The difference between the models grows even larger as the policy age reaches four years and older. The result implies information, not fully captured by the control of observable characteristics.

Formal statistics in table4 also show that there is no robust evidence for informational monopoly in claim severity results. Table4 reports the coefficients of the policy age dummies, the unconditional variance of *ln (severity)*, and σ_{μ}^2, the unobserved heterogeneity. I plugged the lognormal severity as the dependent variable and estimated the coefficients in the linear regression by maximum likelihood method. Unobserved heterogeneity presented by the unconditional variance, σ_s^2 does not differ by specification. This implies that policy age variables do not capture any information not fully explained by observable characteristics. A random effect by policyholder also seems to exist but there is no clear difference in size among specifications and the size itself is almost zero. This partly comes from the fact that the number of drivers who get involved in accidents more than once is relatively small to generate any random effect. Measures of the goodness of fit of the model are confusing. Comparing specification (A), which includes a full set of dummy variables for policy age and specification (B) including only a policy age 0 dummy, the log-likelihood ratio is 0.96 where the five percent chi-square distribution is 9.48. Thus, specification (A) does not raise the goodness of fit of the model. The results from AIC and BIC are consistent with log-likelihood results. Further, coefficients are not monotonically related to policy age. Comparing model B and C points out no significance of the dummy variable for policy age 0. In short, the inclusion of policy age 0 dummy variable does not increase the goodness of fit of the model. In short, it is difficult to confirm the existence of an informational monopoly in the claim severity analysis.

Unlike the results on claim frequency risk, it is impossible to find a monotonic relationship between policy age and claim severity risk after controlling for various control variables.

Until now, I confirmed the existence of an informational monopoly in the Korean auto insurance market, due to the low average claim frequency risk of older policies.

Policy Age, Premiums, and Profitability

Another hypothesis from asymmetric learning identified in this article is that profitability decreases with policy age at a low rate. Recall the unconditional loss-ratio presented in figure1. The figure is consistent with theory and now I explore whether the relationship holds after controlling observable characteristics. Table5 shows the relationship between policy age and charged premiums. Coefficients are from ordinary least squares (OLS) regressions, where the dependent variable is the log of charged premiums. Specifications (A), (B), and (C) are identical to former sections. Results show evidence for a monotonic relationship between policy age and profitability after controlling for all observable variables. Coefficients decrease as policy tenure grows older. Charged premiums for policies that are four years or older fall with a high rate of change. Moreover, the goodness of fit of the models consistently increases as policy age is controlled. Overall, results of absolute measures of profitability cast light on empirical results consistent with asymmetric learning theory. When premium decreases at a slower rate than total claim costs, the insurance firm can make profit. Thus, based on the monotonic trend in premiums by policy age in table5, examination on relative measures are followed to identify whether loss ratios decrease as policy age gets older.

Table6 confirms the results by examining a relative profitability measure, the loss ratio. Loss ratio is generally used to evaluate the profitability of insurance firms, indicating the claim amount to collected premiums. Results are evident; (1-Loss Ratio) is lower for new policies and increases by policy age. Figure4 clearly shows an upward tendency of profitability. To conclude, profitability in both absolute and relative measures provides evidence for strategic behaviors of insurance firms based on their informational advantage compared to rivals.

Policy Age and Claim Frequency: Residual Market

Based on the results of the article, older policies are less risky due to low claim frequency. From this relationship, I conclude the existence of an informational monopoly in the Korean auto insurance market. However, there might be concerns that policy age not only proxies accumulated amount of information gathered by the incumbent insurer so that the negative relationship does not come from the strategic behavior of the insurance company. Considering such remarks, I examine the relationship between claim frequency risk and policy age for contracts that are not voluntarily written by the incumbent insurers but mandatorily assigned to them.

The co-insurance system is equivocal to the residual market in the auto insurance system of the United States. A contract is co-insured when the applicant is too risky and insurers deny writing contracts with. However, to ensure the sustainability of the insurance market, such applicants are assigned to several insurers who share the risk together. Thus, in the case of co-insured contracts, insurers have far less space to strategically control the average risk regardless of the length of policy tenure.

To examine the concept in more detail, I show the claim frequency results for the third-party physical damage (PD) and the first-party physical damage part (PD). I examine the two separately because insurers are forced to accept third-party liability contracts when assigned but have the right to reject first-party physical damage. Based on the fact, I first hypothesize that there will be no linear relationship among policy age and the number of accidents for co-insured polices. Further, controlling for observable characteristics and policy age will be insufficient to explain the actual losses. At last, these two hypotheses will be even stronger for third-party physical damage since the incumbent insurer has no power to control the policies.

Figure5 shows the claim frequency risk of first-party physical damage for co-insured policies. The thin black line (A) indicates the actual claim frequency while (B) and (C) present the mean predicted values of specifications identical to those used in former sections. Same as in figure2, actual claim frequency indicates the mean predicted value from the negative binomial model. Specification (B) includes a dummy for new policies and (C) includes no controls for policy age. First, claim frequency is not monotonically decreasing with policy age. It is noteworthy that both specification (B) and (C) cannot predict actual loss frequency. This confirms the hypotheses stated in this section. Table7 shows the formal statistics of the regressions. After controlling for observable characteristics, there is no tendency by policy age. The goodness of fit does not show any consistent improvements. The dummy variable for new polices has no significant coefficient and the inclusion of the dummy does not improve the fit of the model compared to no control for policy age dummies.

Whereas insurance firms have a certain power to reject a policyholder's application for the first-party PD, the insurer has no choice but to accept the contract when it comes to third-party PD. Figure6 is consistent with the hypothesis that there is no relevance between policy age and claim risk when the insurer has no informational monopoly on the policy. The actual claim frequency marked by the thin black line shows a random plot by policy age. Further, specification (B) and (C) are far from the actual claim frequency. Table8 confirms the results presented in figure6. There is no relationship between policy age and the number of claims after including a full set of control variables. Moreover, comparing the goodness of fit of various specifications shows no consistency to decide which model has stronger explanation power.

Overall, the examination of policies of the residual market supports the fact that decreasing average claim frequency risk by policy age is evidence for informational monopoly and the strategic behavior of incumbent insurers based on informational advantage.

Conclusion

The study was initially motivated by conflicting results in empirical findings on the existence of information monopoly in nonlife insurance contract environment. Kofman and Nini (2013) who examined the Australian auto insurance market concluded that publicly available data capture all information about policyholders, subsequently removing informational monopoly and point out that the NCD rating factor plays an essential role in preventing informational monopoly. The Korean auto insurance market also shares a bonus-malus factor which corresponds to the NCD rating factor in Australia. However, unlike in other countries, the Korean bonus-malus factor is based on accident severity rather than frequency and by sharing an aggregated bonus malus "score", the market is left space for information monopoly since the claim history of a policyholder is not properly reflected. On the other hand, the existence of the Korea Insurance Development Institute may reduce the possibility to accumulate information monopoly since after a certain period of time, all accident details including the nature, time, frequency, and severity accidents is shared by all insurance firms in the market.

Thus, in this article, I examine the existence of an informational monopoly in the Korean auto insurance market. Through finding a positive relationship between policy age and claim frequency after controlling for publicly observable underwriting factors I confirm the existence of informational advantage of incumbent insurers. Further, examination on profitability measures confirms that insurers subsequently generate more profit from older policies. Additionally, results on residual markets imply the existence of a strategic behavior of insurers since claim risk does not systematically decreases with policy tenure when insurers are not able to voluntarily decide whether to write the contract or not.

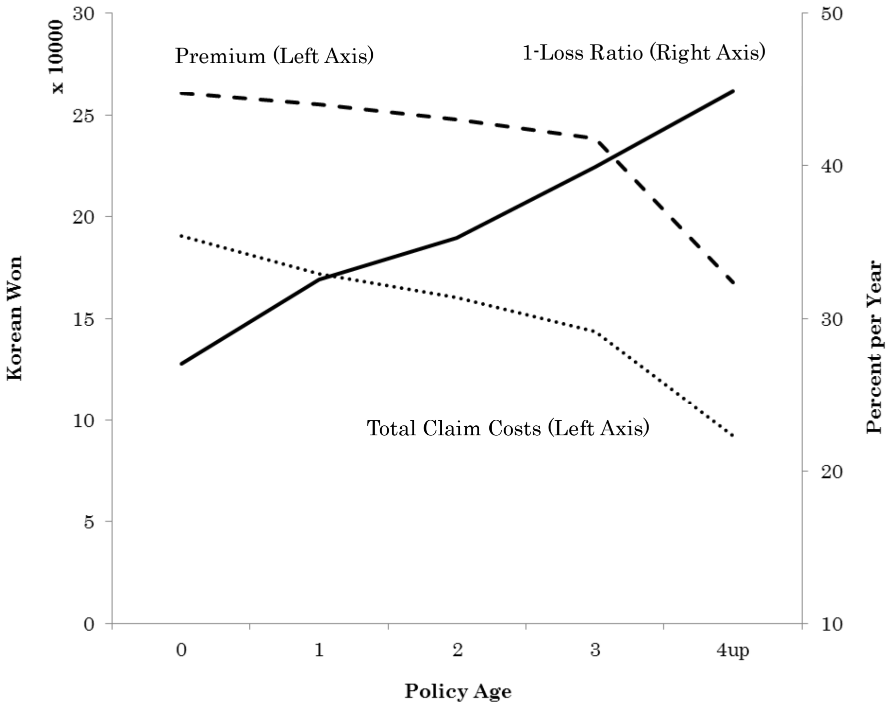
Table 1

Definitions of Variables

Policyholder Characteritics	
Age	Age, in years, of the primary driver of the covered automobile
Male	Dummy variable indicating that the primary driver is a male
Districts	Dummy variables indicating the district of registration
BMS	Bonus Malus coefficient: 11 is the starting class, 1-10 are malue(penalty), and 12-25 are bonus(discount) classes
Automobile Characteristics	
Vehicle age	Age, in years, of the insured automobile
Insured Value	Value, in 100million Korean won, of insured automobile in case of total loss
Capacity	classified into four categories
Foreign	origin of production
Sportscar	1 if car is classified as sportscar
Policy Characteristics	
Policy Age	Age, in years, that the policy has been in-force, 0: less than one year, 7up: more than 7 years
Premium	Total premium charged on the policy
Driving Experience	Driving experience in years, categorized into 8 dummies from 0 to 7 or longer
Age_limit	1 if special contract on age is included(classified into four categories)
Lowmile	1 if special contract on low miles driven is included
Policy performance	
CNT	Number of claims on the policy incurred during the policy period
LOSS	Total value of claims made on the policy during the accident period

Note: This table lists all variables used in the regression models, including all of the control variables

Figure 1
Premium and Costs by Policy Age



Note: Figure1 presents unconditional averages of premium, total claim costs, and loss ratio by policy age. Premium (measured on the left axis) is the annual total amount of premium for a policy. Total claim costs (measured on the left axis) include all payments. The loss ratio is the total losses to total premium, and the figure shows 1-loss ratio (represented on the right axis), which is a common measure of insurers' profitability.

Table 2

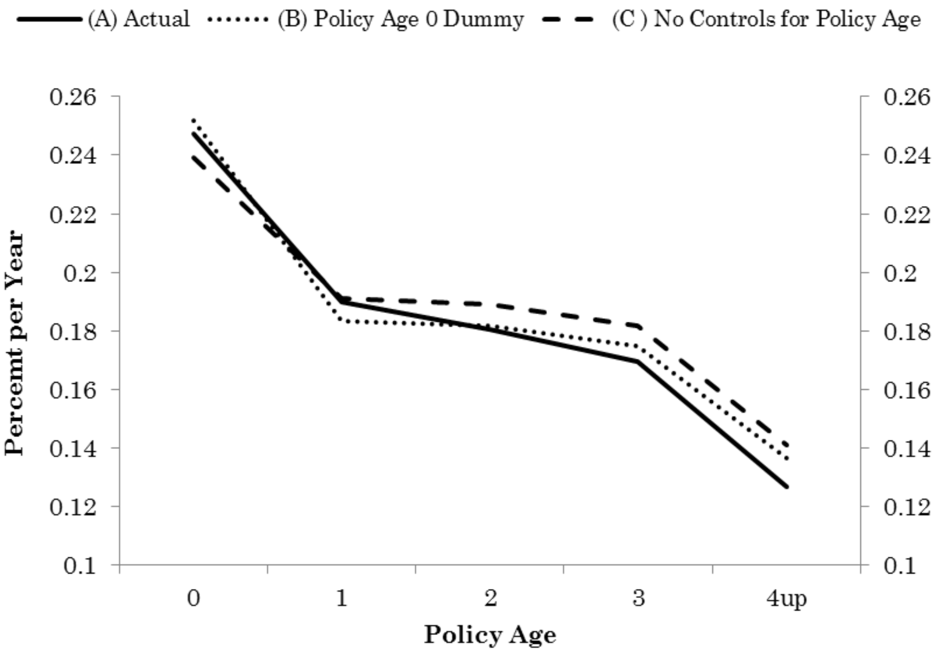
Summary Statistics

Panel A: Distribution of Control Variables							
Variable	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
Age	42.05	13.90	18	31	39	51	80
Male	0.72	0.45	0	0	1	1	1
Vehicle Age	5.10	4.53	0	1	4	9	15
Insured Value	11748447	12159872.98	10000	3890000	8520000	16330000	600800000
Total Premium	255173	280594.41	0	0	219940	369560	11839140

Panel B: Sample Means by Policy Age					
Policy Age	Insured Age	Male	Vehicle Age	Insured Value	BMS
0	39.49	0.72	4.69	12411367.59	12.95
1	40.34	0.72	4.74	12343542.71	13.17
2	42.45	0.72	4.96	11922304.76	13.51
3	45.61	0.70	5.23	11738003.41	13.83
4up	54.80	0.76	7.54	7699884.08	16.09

Note: Table2 presents the summary statistics of policy holder characteristics and samples means by policy age for all the samples.

Figure 2
Predicted Claim Frequency Risk by Policy Age



Note: This figure presents mean predicted annual claim frequencies by policy age for the models summarized in Table3. The thin black line represents the actual claim frequency and corresponds to the predicted values from model (A) which includes a full set of controls for policy age. The dotted line, (B), only includes a dummy variable for the policy age 0. The dashed line represents the predicted value of model C, which included no control variable for policy age.

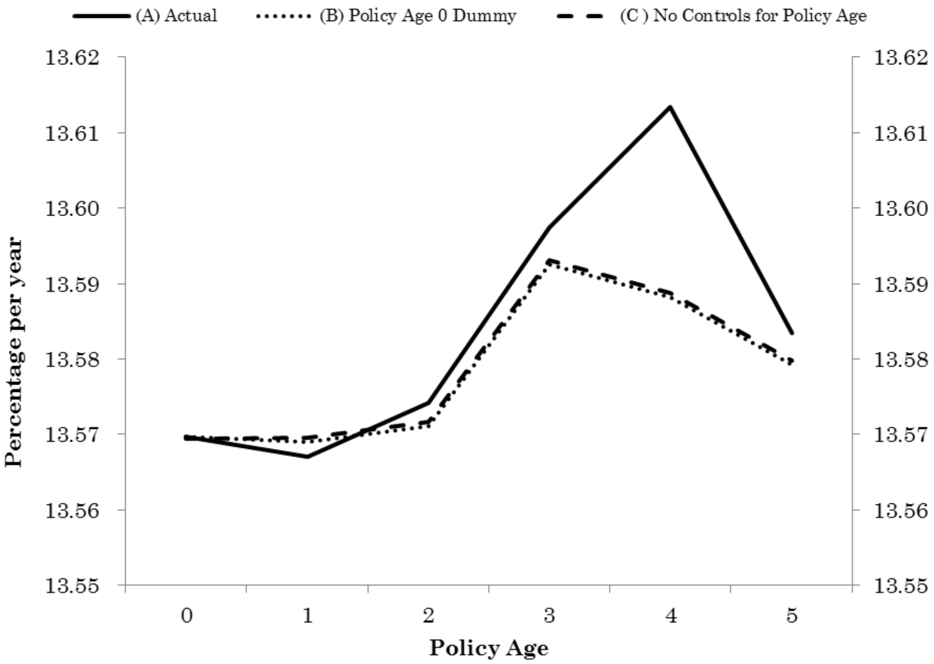
Table 3
Claim Frequency Risk – Impact of Policy Age Variables

	Dependent Variable: Number of Claims		
	(A)	(B)	(C)
Policy Age			
0		0.1022(0.0086)	
1	-0.0676(0.0100)		
2	-0.1094(0.0128)		
3	-0.1341(0.0206)		
4up	-0.1903(0.0149)		
σ_v^2	1.2826(0.0205)	1.2855(0.0205)	1.2911(0.0205)
Log-likelihood	-191325.57	-191359.7763	-191430.0576
AIC	405554.673	405617.0922	405755.6548
BIC	406068.1298	406097.7752	406225.4131

Note: This table presents coefficient estimates and standard errors in parentheses from the negative binomial models for the number of claims. Specification (A) includes four policy age dummy variables for each policy age group; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include control variables listed in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

Figure 3

Predicted Claim Severity Risk by Policy Age



Note: This figure presents mean predicted annual claim severity by policy age for the models summarized in Table4. The thin black line represents the actual claim severity and corresponds to the predicted values from model (A) which includes a full set of controls for policy age. The dotted line, (B), only includes a dummy variable for the policy age 0. The dashed line represents the predicted value of model C, which included no control variable for policy age.

Table 4

Claim Severity Risk-Policy Age Variables

	Dependent Variable: Ln(Claim Severity)		
	(A)	(B)	(C)
Policy Age		0.0087(0.0047)	
0			
1	-0.0028(0.0053)		
2	0.0225(0.0090)		
3	0.0041(0.0166)		
4up	0.0244(0.0383)		
σ_μ^2	0.0036(0.0164)		
σ_s^2	0.0300(0.1409)	0.0292(0.1450)	0.0292(0.1450)
Log-likelihood	0.9744(0.0045)	0.9744(0.0046)	0.9744(0.0046)
AIC	-276480	-276481	-276481
BIC	553066.4	553059.3	553057.4

Note: This table presents coefficients estimates and standard errors in parentheses from OLS regression of claim severity. The dependent variable is the log of claim amount, that is, claim severity. Specification (A) includes 6 policy age dummy variables for each policy age groups; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include all control variables presented in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

Table 5

Charged Premiums-Policy Age Variables

	Dependent Variable: Ln(Premium)		
	(A)	(B)	(C)
Policy Age			
0		0.00013(0.0012)	
1	-0.00749(0.0013)		
2	-0.00231(0.0020)		
3	0.00001(0.0041)		
4up	0.04477(0.0025)		
Log-likelihood	-58125	-58344	-58344
AIC	116351.7	116784.6	116782.7
BIC	116889.3	117290.6	117278.1

Note: This table presents coefficients and standard errors in parentheses from the OLS models. The dependent variable is the log of the charged premiums. Specification (A) includes four policy age dummy variables for each policy age groups; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include all control variables presented in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

Table 6

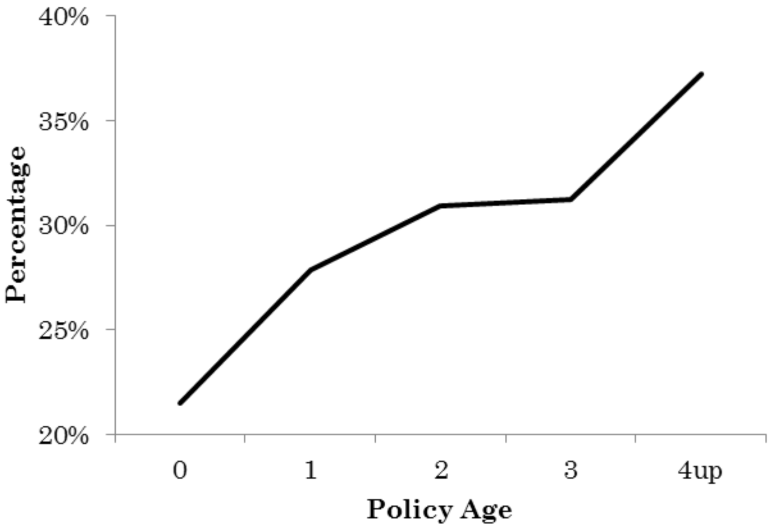
(1-Loss Ratio)-Policy Age Variables

	Dependent Variable: (1-Loss Ratio)		
	(A)	(B)	(C)
Policy Age			
0		-0.0899(0.0097)	
1	0.0567(0.1090)		
2	0.0936(0.0159)		
3	0.0859(0.0345)		
4up	0.1925(0.0183)		
Log-likelihood	-651583	-651605	-652116
AIC	1303269	1303306	1303390
BIC	1303806	1303812	1303886

Note: This table presents coefficients and standard errors in parentheses from the OLS models. The dependent variable is (1-Loss Ratio), a relative measure of profitability. Specification (A) includes four policy age dummy variables for each policy age groups; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include all control variables presented in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

Figure 4

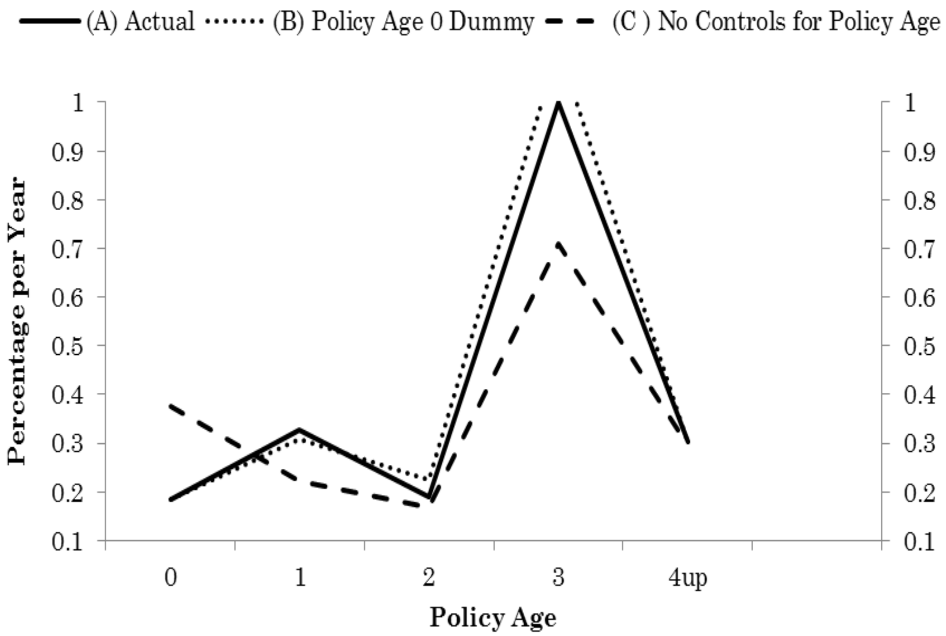
(1-Loss Ratio)-Policy Age Variables



Note: This figure presents (1-Loss Ratio), a relative measure of profitability by policy age for the models summarized in Table6. The thin black line represents the actual (1-Loss Ratio) and corresponds to the predicted values from model (A) which includes a full set of controls for policy age.

Figure 5

Claim Frequency Risk-Impact of Policy Age_First-Party PD in the residual market



Note: This figure presents mean predicted annual claim frequencies by policy age for the models summarized in Table7. The thin black line represents the actual claim frequency and corresponds to the predicted values from model (A) which includes a full set of controls for policy age. The dotted line, (B), only includes a dummy variable for the policy age 0. The dashed line represents the predicted value of model C, which included no control variable for policy age.

Table 7

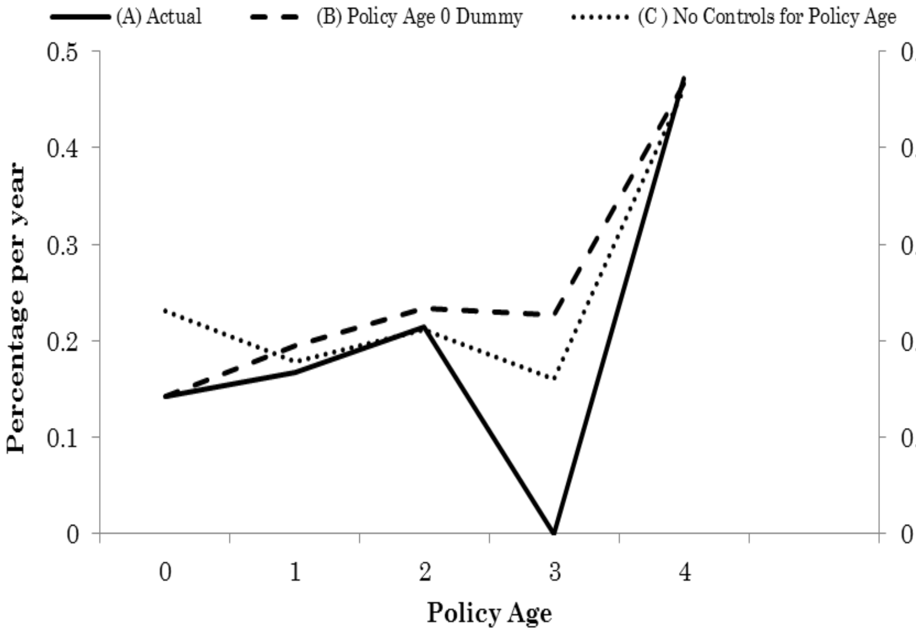
Claim Frequency Risk-Impact of Policy Age_First-Party PD in the residual market

	Dependent Variable: Number of Claims		
	(A)	(B)	(C)
Policy Age			
0		-1.3365(0.764)	
1	1.4315(0.908)		
2	1.1582(0.984)		
3	1.1890(1.561)		
4up	1.4902(0.939)		
σ_v^2	0.529(0.383)	0.51(0.374)	0.597(0.394)
Log-likelihood	-122.73	-122.8073	-124.5859
AIC	373.599	367.7512	369.3084
BIC	531.8844	515.4842	513.524

Note: This table presents coefficients estimates and standard errors in parentheses from the negative binomial models for the number of claims for First-party PD, co-insurance policies. Specification (A) includes 6 policy age dummy variables for each policy age groups; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include all control variables presented in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

Figure 6

Claim Frequency Risk-Impact of Policy Age_Third-Party PD in the residual market



Note: This figure presents mean predicted annual claim frequencies by policy age for the models summarized in Table8. The thin black line represents the actual claim frequency and corresponds to the predicted values from model (A) which includes a full set of controls for policy age. The dotted line, (B), only includes a dummy variable for the policy age 0. The dashed line represents the predicted value of model C, which included no control variable for policy age.

Table 8

Claim Frequency Risk-Impact of Policy Age_Third-Party PD in the residual market

	Dependent Variable: Number of Claims		
	(A)	(B)	(C)
Policy Age			
0		-0.657(0.665)	
1	0.5189(0.868)		
2	0.5794(0.881)		
3	-21.9964(1575)		
4up	0.8157(0.7224)		
σ_v^2	0.1024(0.198)	0.1066(0.199)	0.1124(0.202)
Log-likelihood	-151.17	-151.53	-152.08
AIC	445.0144	439.7432	438.8464
BIC	603.2997	587.4763	583.062

Note: This table presents coefficients estimates and standard errors in parentheses from the negative binomial models for the number of claims for First-party PD, co-insurance policies. Specification (A) includes 6 policy age dummy variables for each policy age groups; specification (B) includes only control for new policies, and (C) includes no control variable for policy age dummies. All models include all control variables presented in table2. AIC refers to the Akaike information criterion, and BIC is the Bayesian information criterion.

References

Chiappori, P.-A., and B. Salani´e, 2000, Testing for Asymmetric Information in Insurance Markets, Journal of Political Economy, 108: 56-78.

Cohen, A., 2012, Asymmetric Learning in Repeated Contracting: An Empirical Study, The Review of Economics and Statistics, MIT Press, 94(2): 419-432

Cohen, A., and P. Siegelman, 2010, Testing for Adverse Selection in Insurance Markets, Journal of Risk and Insurance, 77(1): 39-84.

D'Arcy, S. P., and N. A. Doherty, 1990, Adverse Selection, Private Information, and Lowballing in Insurance Markets, Journal of Business, 63: 145-164.

Dionne, G., and N. A. Doherty, 1994, Adverse Selection, Commitment, and Renegotiation: Extensions to and Evidence From Insurance Markets, Journal of Political Economy, 102(2): 209-235.

Kofman, P., Jullien B., Salani´e, B., and Salani´e, F., 2006, Asymmetric information in insurance: general testable implications, RAND Journal of Economics, 37(4), 783-798.

Kofman, P., and G.P. Nini, 2013, Do Insurance Companies Possess an Informational Monopoly? Empirical Evidence from Auto Insurance, Journal of Risk and Insurance, 80(4): 1001-1026.

Kunreuther, H., and M. Pauly, 1985, Market Equilibrium With Private Knowledge: An Insurance Example, Journal of Public Economics, 26: 269-288.

국문초록

본 연구는 한국 대형 보험사의 데이터를 이용하여 가입자의 평균 위험 정도와 가입기간의 관계를 살펴본다. 가입기간이 길어질수록 가입자의 평균 위험도는 낮아진다. 이는 가입기간이 길어질수록 사고의 빈도가 줄어드는 현상에 기인한다. 나아가 가입기간이 증가함에 따라 이들 가입자로부터 얻는 보험사의 이윤이 증가하는데 이러한 연구결과는 비대칭적 학습에 대한 이론의 예측과 일치하며 한국 자동차보험 시장에서의 정보 독점이 일정 정도 존재함을 지지한다. 나아가 공동인수 대상 가입자들에 대한 연구는 가입자 정보에 기반한 보험사들의 전략 활용을 방증한다.

주요어: 비대칭적 학습, 정보 독점, 자동차보험, 공동인수

학번: 2014-20460