Testing for Asymmetric Information and Learning

-Evidence from the Korean Auto Insurance Market-

Abstract

This article draws on a set of data from the Korean automobile insurance industry to investigate the presence of asymmetric information and learning. I find that there exists a conditional correlation between the choice of Collision and Comprehensive coverage and the probability of accidents. This finding implies that the better insured actually cause more claims, although the difference in the number of claims seems less significant in economic contexts. I confirm that the degree of information asymmetry is more significant in contracts with new policyholders, young drivers with less than 3 years of driving experience, and senior drivers with more than 7 years of experience. Tests using premiums provide results consistent with tests considering variables related to the characteristics of policyholders, characteristics of their vehicles, and the contract choices they make.

Keywords: adverse selection, asymmetric learning, auto insurance, insurance premium

I. Introduction

The existence of information asymmetry between an insurer and an insured has long been considered a main impediment to finding equilibrium in insurance markets. According to adverse selection theory, since people with higher risk tend to further extend their coverage, there exists a positive correlation between the choice of better coverage and a higher probability of accidents. Ever since Rothschild and Stiglitz (1976) established that information asymmetry may lead to inefficient outcomes and market failure in competitive insurance markets, economic theorists have tried to develop a model to test for the presence of adverse selection.

This positive correlation can be explained through the concept of moral hazard: agents with more comprehensive insurance will have less incentive to take preventive measures, which in turn leads to a higher probability of accidents. The moral hazard concept has the same results as adverse selection, although the causality is reversed. In contrast, other models yielding opposite predictions to those of Rothschild and Stiglitz (1976) do exist. One is the concept of advantageous selection, which states that people who have more insurance are actually at lower risk, because they have lower risk tolerance and attentive in avoiding accidents. According to this argument, there should be a negative correlation between insurance coverage and accident probability.

The first purpose of this article is to test for the presence of adverse selection in the Korean automobile insurance market. Specifically, I seek to confirm if there is a conditional correlation between the choice of Collision and Comprehensive coverage and the number of claims reported by using Korean insurance data to repeat tests conducted by Chiappori and Salanié (2000). I only consider claims for Bodily Injury I and Property Damage coverage, because these can be observed for all policyholders.

Although I do not exclude the possibility that any positive correlation between coverage choice and risk occurrence is due to moral hazard, I have elected to not concentrate on distinguishing the two concepts. Additionally, my focus is only on the number of claims, because the amount of loss is generally considered to be more related to ex-post moral hazard, which means that policyholders with better coverage tend to claim indemnity in larger amounts when involved in accidents.

Another purpose of this article is to see whether people with greater driving experience have learned more about their type and degree of risk, in which case the problem of adverse selection would be stronger for those individuals. Cohen (2005) shows that there exists asymmetric learning as well as asymmetric information between insurer and insured in the Israeli car insurance market. He establishes that the informational advantages of insureds were acquired from their driving experience and shows that the correlation between choice of deductible and number of claims is more correlated for people who have been driving longer.

In addition, it is also assumed that insurers can also learn about their client's risk. Cohen (2012) shows that the ratio of premium to loss is larger for repeat customers who have a good claim history and renew their contract with the same insurer. Han (2016) confirms similar phenomena in the Korean auto insurance market. For this reason, I consider the impact of policy age on the degree of adverse selection.

The final goal of the article is to determine if premium information is enough to examine the presence of adverse selection. Zavadil (2015) suggests the model in which all the characteristic variables are replaced by a single-variablepremium. He proves that this new method produces the same results as those from the non-parametric models suggested by Chiappori and Salanié (2000). I extend the application of his idea to parametric models.

There are several articles testing for adverse selection in the Korean insurance market, but some of the variables are inappropriately selected, and the regression models have restrictions. Therefore, this paper would be the first to follow the procedures of Chiappori and Salanié (2000) and to examine the effectiveness of the premium-based model in the Korean automobile insurance market.

The rest of the article is organized as follows. The next section reviews previous studies related to information asymmetry in detail. Section 3 covers background information about the Korean auto insurance market and data. Section 4 presents empirical methods and results. The final section concludes the article.

II. Hypothesis Development

A. Related Literature

Earlier empirical studies, such as Puelz and Snow (1994), seem to support the existence of adverse selection in the automobile insurance market. However, subsequent studies question the results of their analysis. Chiappori and Salanié (2000) criticize Puelz and Snow's (1994) use of both linear functional forms and only about 20 variables, which might create a spurious conclusion. To solve the limitations of previous literature, they have adopted new empirical methods. Two parametric models and three non-parametric models are presented in the paper, all of which produce the same results. This method is now considered as a standard procedure to test adverse selection. Their analysis is focused on young drivers who have driven for less than 3 years, a focus that excludes the impact of experience rating. It has been shown that there is no evidence of adverse selection for young drivers in the French automobile insurance market.

On the other hand, Cohen (2005) proves the existence of asymmetric information in the Israeli auto insurance market. He tests the presence of adverse selection in the pool of new customers, who are most likely to have an informational advantage over an insurer, and finds that new customers with a low deductible are associated with more accidents and higher total losses for the insurer. It is also suggested that the extent of adverse selection is related to an insured's driving experience. He found that there was evidence of adverse selection for drivers new to the company with more than 3 years of driving experience but not for those with less than 3 years of driving experience. This result is consistent with Chiappori and Salanié. Also, it means that the insured have learned some private information about their risk from greater driving experience.

However, Cohen (2005) could not control for a policyholder's past claim history, a factor which likely affects both price and selection of subsequent contracts. He used data from the Israeli market, where insurers are not required to share information about their former customers with other companies. As a result, the information an insurer receives about a new policyholder mainly depends on that customer's self-reporting. In contrast, the Korean market has an information sharing system that offers insurance companies free and full access to a national database of policyholders' complete records on accident claims. Therefore, it is assumed that an experience-based rating system based on policyholders' past claim history would reduce any effects gained through information an agent has otherwise acquired.

From a similar point of view, some studies investigate the existence of asymmetric learning by an insurance company from repeated contracts. Cohen (2012) shows that an insurer makes higher profits in transactions with repeat customers who have a good claim history with that insurer. This is because the insurer reduces their premium by a lesser degree than the degree of reduction in expected costs they experience. In other words, the insurer knows better than the insured about their risk and uses this informational advantage to determine their price strategy. In contrast, it is hard to confirm asymmetric learning from repeat customers with bad claim histories, because they are more likely to switch to other insurers to avoid high premiums. Han (2016) examines the presence of an informational monopoly in the incumbent insurer with repeat customers in the Korean auto insurance market. This study found that there is a positive correlation between claim frequency of policyholders and their policy age; additionally, it appears that the company gains more profits form older policies.

Zavadil (2015) confirms that when including an experience rating into the set of conditional variables, evidence for information asymmetry vanishes in the Dutch car insurance market. This result is mainly due to the fact that experience ratings include all past claims, which allows insurers to better price individual risk. Since the Korean auto insurance industry also has a similar system, such ratings are expected to alleviate the problem of information asymmetry. The more important contribution of this article is that it presents a new, premium-based model to test for adverse selection. The author argues that insurance premiums reflect policyholders' underlying risk in the best possible way. Insurers calculate premiums using a complex pricing model that considers multiple risk factors and examine the effectiveness of their models periodically. The premium-based model brings the same results as the non-parametric models suggested by Chiappori and Salanié (2000).

There are not many Korean studies related to this issue because it is hard to obtain individual-level panel data. Roh and Ha (2008) investigate the correlation between accident occurrence and coverage choices. They used a logit model with four different dependent variables: the coverage choices of Bodily Injury II, Property Damage, Medical Payment, and Collision and Comprehensive, all of which were optional in 2008. They estimated expected accident probability for each policyholder and defined it as a dependent variable. Although their empirical method is very similar to Dionne et al. (1997), it has some drawbacks. First, including information about the driver's region in control variables is inappropriate because it is illegal to use them as a pricing factor in the Korean insurance market. Second, using expected accident probability as a proxy for risk occurrence could produce measurement errors.

Kim (2009) investigates the evidence of adverse selection in the Korean market by examining the conditional correlation between the number of claims and the choice of low deductible and full coverage, but this method also has flaws. Above all, it uses claim frequency in year *t-1*, while the choice of policy options occurred in year *t*. This approach is not valid for testing the presence of adverse selection because there is no asymmetric information. Since both the insurer and the insured know about the insured's claim history, it would be applied to the next year's premium, which would also affect the insured's choice of policy options. Further, since there were very limited options for deductibles until 2009, it would not give any significant information on an insured's risk type. In addition, Kim (2009), unlike Cohen (2005), shows that the extent of adverse selection is negatively related to years of driving experience in the Korean market.

B. Predictions

To test for the presence of adverse selection in the Korean automobile insurance market, I have developed four testable predictions.

H1: There exists a conditional correlation between the number of accidents and the choice of Collision and Comprehensive coverage.

The first hypothesis is to test for a conditional correlation between the choice of Collision and Comprehensive coverage (CNC) and the number of claims reported. This hypothesis predicts that people with CNC coverage have more accidents with third-party damage, which is inconsistent with Chiappori and Salanié (2000) and Zavadil (2015). Unlike these two previous articles, I consider both young and senior drivers. Also, I presume that premise, considering that:

H2: The degree of conditional correlation is stronger for the insured with longer years of driving experience.

My second hypothesis examines whether or not there exists asymmetric learning depending on an insured's driving experience. As Cohen (2005) shows, I assume that policyholders do learn about their risk from their driving experience, in which case adverse selection is more strongly observed in the insured with more years of driving experience.

H3: The degree of conditional correlation is weaker for the insured with an *older policy age.*

There can be asymmetric learning by the incumbent insurer through repeated contracts. Consistent with the results of Han (2016), I presume both that an insured with an older policy age will have fewer claims than others and that the degree of asymmetric information will be the strongest for newcomers to a company. Policy age indicates the number of years that the insured has been under contract with the incumbent insurer.

H4: Tests conducted with premium information only instead of all exogenous variables will produce the same results.

The final hypothesis is to examine whether premium information is enough to test for adverse selection. Zavadil (2015) uses Dutch auto insurance data to confirm that tests based on premium produce results consistent with conventional tests. As I use data from the largest and most profitable auto insurer in Korea in this study, I predict that its premiums well reflect all exogenous policyholder variables related to their degree of risk.

III. Background Information and Data

A. Car Insurance in Korea

In the Korean automobile insurance market, all auto insurance policies are in effect for one year. Customers can either change their insurer or renew their policy each year, and any changes result in the recalculation of their premiums.

Insurance companies generally offer 6 types of coverage: Bodily Injury I (BI-I), Bodily Injury II (BI-II), Property Damage (PD), Collision and Comprehensive (CNC), Medical Payment (MED), and Uninsured Motorist (UM). The first three cover the other driver's and passengers' medical bills and any damages for which the insured driver is deemed responsible. Bodily Injury I and Property Damage are called "liability coverage," which are legally required, and the rest is optional. Collision and Comprehensive pays the repair of the insured's vehicle, while Medical Payment pays for an insured's medical care of any kind when he/she is in an accident and Uninsured Motorist pays if he/she gets involved in an accident caused by uninsured drivers.

Of the optional coverage options, CNC has caused insurers issues in recent years because of the high degree of information asymmetry between insured and insurer. Before 2010, only one type of CNC coverage was available, and it had a single loss limit and a single deductible. Any accidents exceeding the deductible would be indemnified by insurers. In 2010, they changed the coverage options to choose the upper loss limit among four options provided. This change made both to attract more customers and to better classify their risk type. However, counter to expectations, the rate of loss surged to 92%, an increase mainly due to ex-post moral hazard. In response, they implemented a new policy which a single deductible with flexible deductibles that vary depending on the choice of upper loss limit. Additionally, these deductibles indicate the amount of money that an insured should pay, which is 20% of the loss amount. Since the Korean car insurance industry's deductible system is much more complicated than that of other markets, it is difficult to use it to test for adverse selection.

Korean insurance companies use the Bonus-Malus (BM) experience rating system. At the beginning of the contract period, each policyholder is assigned to a certain BM class that reflects his past claim history. In 2012, there were 24 classes, and each class had a fixed rate premium. All new policyholders begin in class 11Z, and 24P is the highest class. The Korean BM scheme is unique, because it reflects claim severity as well as claim frequency, while the BM system used in other countries only considers claim frequency.

[Table 1 about here]

Table 1 shows how to surcharge or discount premiums. Depending on the number and type of claims occurring during the past contract year (generally one year), each client gets points that will affect which BM class he/she will be placed in during the next contract period. If an insured has had no at-fault claims for the past 3 years and his/her BM class is no lower than 11Z, he/she can move up one class, and the premium will be discounted accordingly. If his/her BM class is lower than 11Z, which means he/she is paying a surcharged premium, he/she can move to class 11Z after 3 years without any at-fault claims.

B. Sample Selection

This paper is based on data sourced from one of the largest automobile insurance companies in Korea. It includes all contracts executed during 2012. The raw data contains 5,543,856 policies. Of these, I excluded 3,750 co-insured contracts and any policies starting after March; this leaves 1,273,527 observations. By restricting the starting point, I can avoid heterogeneity problems caused by special events in the market, such as premium fluctuations or regulatory changes. The size of the remaining data set was still large because the company renewed policies even if there were any small changes. For this reason, I first merged the data for the insured who had updated their policies. Then, I selected contracts which have been in force for at least one full year. By matching the contract period of policyholders, it became more convenient to run empirical regressions. Of the remaining 620,715 contracts, I deleted contracts which have missing values. The final data set consists of 584,778 contracts.

C. Summary Statistics

Table 2 presents descriptive statistics of policies with and without CNC coverage. Age means the age of primary driver of the insured car. Gender indicates 1 if the driver is male. The Bonus-Malus coefficient has a value ranging from 1 to 24. Driving experience is expressed in years since the insured started to drive, with values from 0 to 7, where 0 indicates less than a year of driving experience and 7 indicates more than seven years of experience. About 68% of policyholders in the data set have been driving for more than seven years. Policy age has a value from 0 to 4, which represents the number of years since the insured entered into a contract with the incumbent insurer. Since I have individual panel data from 2009 to 2012, I have compared the list of policies by year and have assigned a policy age to the 2012 selected sample.

[Table 2 about here]

It seems that policyholder characteristics of two groups are similar. We can see that people with CNC coverage are younger and more likely to be female. In contrast, there are big differences in car characteristics between two subsamples. Car age is expressed in years, which represents the period of use. Car type is classified into two groups depending on capacity. Cars which can carry fewer than seven people are divided again into four types, with values from 1 to 4. The smaller the car type value is, the smaller the engine size. High-occupancy vehicles have a value of 5. People with newer, bigger, more expensive, and foreign cars are more likely to buy CNC coverage. This result is logical because the choice of CNC coverage and its associated premium are reflected by the expected repair cost of the insured car. Some differences are also observed between groups in regard to policies. It is interesting that the premium for the group with CNC is lower, while their number of claims and the loss amount are actually larger than the group without CNC. People in the first group (with CNC) have 0.15 average claims for BI-I and PD, which is approximately 3.4% more in claims than the other group (without CNC). Their average amount of loss per claim is also 7.7% higher, while their premiums are 3.1% lower.

IV. Empirical Analysis

In this section, I explain empirical methods and results. First, I examine the existence of adverse selection using two parametric methods suggested by Chiappori and Salanié (2000). Then, I evaluate whether there is asymmetric learning depending on driving experience or policy age, following Cohen (2005). To prove the last hypothesis, I run the same regressions with the premium variable instead of the set of exogenous variables.

A. The Presence of Asymmetric Information

To test for the existence of adverse selection, I directly borrow the model from Chiappori and Salanié (2000). Let $\mathrm{i} = 1, \ldots, \mathrm{n}$ denote individuals and define two dependent variables y, z and a set of control variables X as below.

$y_i = 1$ if agent i bought CNC coverage and $y_i = 0$ if not.

$z_i = 1$ if agent i had a claim at fault with third-party damage and $z_i = 0$ if not.

X_i: a set of exogenous variables for agent i.

[Table 3 about here]

Specific details about the variables are thoroughly explained in Table 3.

Accident occurrence is defined as the number of claims for BI-I and PD, which are at-fault claims with third-party damage and are observed for all agents. X_i's (independent variables in Table 3) are variables for characteristics of agent i, his/her car, and policy. Policyholder's age is categorized into 7 dummies. Except for the value of the car and premium, all other variables are dummy and constant. 40 control variables in total are used in the following probit regressions.

First, two probit models can be formed, one for the choice of coverage and the other for accident occurrence:

$$y_i = 1(X_i\beta + \epsilon_i > 0)$$
$$z_i = 1(X_i\gamma + \eta_i > 0).$$

Chiappori and Salanié (2000) first estimate these two probits independently, weighing each agent by the number of days under insurance, and then compute the generalized residuals $\hat{\epsilon}_i$ and $\hat{\eta}_i$ as follows:

$$\hat{\epsilon}_i = E(\epsilon_i|y_i) = \frac{\phi(X_i\beta)}{\Phi(X_i\beta)}y_i - \frac{\phi(X_i\beta)}{\Phi(-X_i\beta)}(1-y_i),$$

where ϕ and Φ denotes the density and cumulative distribution function of $N(0,1)$. The test statistic is defined by

$$W = \frac{(\sum_{i=1}^{n} \omega_i \, \hat{\epsilon}_i \hat{\eta}_i)^2}{\left(\sum_{i=1}^{n} \omega_i^2 \ \hat{\epsilon}_i^{\ 2} \, \hat{\eta}_i^{\ 2}\right)}$$

$$W \sim \chi^2(1) \ \ if \ \ cov(\epsilon_i, \eta_i) = 0,$$

where ω_i is the length of time covered by the policy. Since I use only samples with one full contract year, ω_i is 1 for all agents.

Second, in case the two probits are dependent, it is also suggested to estimate a bivariate probit in which $\epsilon_i, \eta_i \sim N(0,1)$ with a correlation coefficient ρ, then test the null $\rho = 0$ and get a confidence interval for ρ.

[Table 4 about here]

Table 4 reports the result of the probit regressions. The first test gives a W statistics equal to 20.665, so the null of conditional independence is strictly rejected. The bivariate probit also rejects the null and gives an estimate of correlation coefficient ρ equal to 0.028 with a standard error of 0.006.

I conduct the same probit regressions using only premium instead of all independent variables to see whether the premium contains enough information about the agent's risk. The results from tests with just the premium are consistent with tests with conditional variables. It gives a W statistics equal to 48.396 and an estimate of correlation coefficient ρ equal to 0.039 with a standard error of 0.006.

The results from the probit regressions are inconsistent with Zavadil (2015), in which including BM class as a control variable removes any evidence of information asymmetry. This is probably because, unlike the Dutch automobile insurance market, the Korean Bonus-Malus system is also based on claim severity. In order to compare more precisely with Chiappori and Salanié (2000), I also repeat the regression models with young policyholders who have driven for less than 3 years, and the results again reject the null. In conclusion, people in the Korean auto insurance market with CNC coverage tend to have more liability coverage claims.

B. Asymmetric Learning on Private Information

For examining the second and third hypothesis, I use ordinary least squares (OLS) and Poisson specifications as Cohen (2005) does. These methods are helpful in comparing the degree of asymmetric information depending on driving experience and policy age.

To see if policyholders learn about their risk from their driving experience, I first divide the whole sample into eight groups (Group0 to Group7+) according to the number of years of driving experience; people in Group0 have less than one year of driving experience, while people in Group7+ have more than seven years of experience. About 68% of policyholders are included in Group7+. Then, I regress the number of claims for BI-I and PD on a dummy variable indicating whether CNC coverage was chosen or not.

[Table 5 about here]

In Table 5, I test four regression models in total: two in OLS and two in Poisson. Models (1) and (3) include the set of independent variables shown in Table 3, save for EX0 to EX5. On the other hand, models (2) and (4) only use the premium for conditioning as a control variable. When comparing the outcomes of regressions with the whole sample, all four models show that there exists a positive and significant correlation between the number of claims and coverage choice. Additionally, using only premium information makes no difference in the outcome. However, it can be inferred that the degree of information asymmetry is economically insignificant but statistically significant. For the better insured, the probability of having at least one accident is higher by 0.8% in OLS and higher by 6% in Poisson regressions. The gap in accident probability seems negligible in comparison with Cohen (2005), in which the probability is higher by 4% in OLS, 16% in Poisson regressions.

When seeing the results of each group, models (1) and (3) show that there exists a positive correlation between claim frequency and coverage. In particular, the coefficients for the insured who drive fewer than two years or more than seven years are significant at the 1% level, which is not consistent with Cohen (2005). Cohen (2005) suggests that the coverage-claim correlation seems significant for new policyholders with more than three years of driving experience. However, in the Korean auto insurance market, no distinct evidence is observed for proving asymmetric learning by the insured. The most interesting outcome is the negative correlation between coverage choice and accidents observed in both Group3 and Group4 when only considering premium information. This means that policyholders in such groups tend to pay high premiums for their actual risk.

To investigate the presence of an information monopoly in the incumbent insurer, I divide the whole sample into 5 groups (Group1 to Group4+) according to policy age. People in Group0 are newcomers, and people in Group4+ have maintained their contracts for more than four years. About 30% of the sample are included in Group0, and 18% are in Group4+. I then run four regression models in the same way as in Table 5. In this time, variables named Policy age0 to Policy age3 are excluded, and EX0 to EX5 are included in the control variables.

Before performing a regression analysis, it was assumed that the correlation between policy age and claim frequency might be weaker for older policies due to the insurer's information monopoly. Han (2016) proves the existence of a negative correlation between claim frequency and policy age in the Korean automobile insurance market. Results from models (1) and (3) seem to accord with this prediction, because coefficients are insignificant in Group4+. However, the outcome from models (2) and (4) is not consistent with the prediction. Coefficients are significant except for Group2. It can be inferred that the coverage-claim correlation does not vanish through repeated contracts.

[Table 6 about here]

It seems that these results are derived from the choice of the independent variable and the characteristics of the data used in this article. Figure 1 shows the average value of CNC, the number of claims, and premiums depending on policy age. All values are standardized by the values in Group0. The older the policy is, the lower the probability of having both claims and CNC coverage. Older policies also pay lower premiums, which may lead to the results seen in Table 6.

[Figure 1 about here]

V. Conclusion

This article examines the presence of asymmetric information and learning in the Korean automobile insurance market. I have built four hypotheses with a sample of 584,778 individual level contracts in 2012. The first hypothesis is to investigate if there is a conditional correlation between the choice of Collision and Comprehensive coverage and the number of claims. The second verifies if the insured learns about their risk through their driving experience. In contrast, the third hypothesis confirms the existence of an information monopoly on the part of the incumbent insurer through repeated contracts. The final hypothesis tests for the effectiveness of the premium-based model used by Zavadil (2015). When calculating premiums and the number of claims, I only consider coverage for Bodily Injury I and Property Damage accidents, as these can be observed for all policyholders.

Using the probit regressions suggested by Chiappori and Salanié (2000), I find that there is a conditional correlation between the choice of CNC coverage and claim frequency, even after controlling for the effects of exogenous variables, such as the characteristics of policyholders, characteristics of their vehicles, and their contract options. It means that policyholders who buy CNC coverage are more likely to have an at-fault accident with third-party damage. Also, the results confirm that the degree of information asymmetry is significant in contracts with newcomers, young drivers with less than 3 years of driving experience, and senior drivers with more than 7 years of driving experience. However, since this article fails to find systematic impacts of driving experience and driving experience on the degree of asymmetric information, it suggests that there is no apparent evidence of asymmetric learning in the Korean auto insurance market.

Further, it can be inferred that premiums do not reflect this difference in accident probability, because models with premium information also reject the null of coverage-accident independence in all regression analysis. It indicates that while people with CNC coverage have a higher risk than those who only purchase liability coverage, they tend to pay lower premiums on average. There are three possible interpretations of these results.

First of all, the presence of adverse selection in CNC coverage can be due to the pricing system. The company probably recognizes that people who buy more insurance are riskier. Nevertheless, it can be regarded as an unfair response to charge higher premiums on liability coverage to people purchasing optional coverage. Instead, the insurer might charge a relatively high premium on a CNC to adjust the total loss ratio. The actual loss ratio of CNC in 2012 was 64.8%, which is low in comparison with that of BI-I and PD (about 80%).

Second, as mentioned above, the bonus-malus system in the Korean auto insurance market is based more on claim severity than on claim frequency. This may prevent the insurer from assigning proper premiums relative to risk. Regulators have noticed the flaws of this system for a few years now and have been trying to find alternatives.

The last possible answer is that the degree of information asymmetry is likely too small to affect a company's profitability, so insurance companies do not have a strong impetus to remove it. The probability of a better insured individual having at least one accident is 0.8% higher based on OLS and 6% higher based on Poisson regressions than for the less insured. This gap does not seem substantial in comparison with the results of Cohen (2005).

References

Chiappori, Pierre-André, and Bernard Salanié, 2000, Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy*, 108(1): 56-78.

Cohen, Alma, 2005, Asymmetric Information and Learning: Evidence from the Automobile Insurance Market, *The Review of Economics and Statistics*, 87(2): 197-207.

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

Dionne, Georges, Christian Gouriéroux, and Charles Vanasse, 2001, Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment, *Journal of Political Economy*, 9(2): 444-453.

Puelz, Robert, and Arthur Snow, 1994, Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market, *Journal of Political Economy*, 102(2): 236-257.

Rothschild, Michael, and Joseph E. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics*, 90(4): 629-649.

Zavadil, Tibor, 2009, Dynamic Econometric Analysis of Insurance Markets

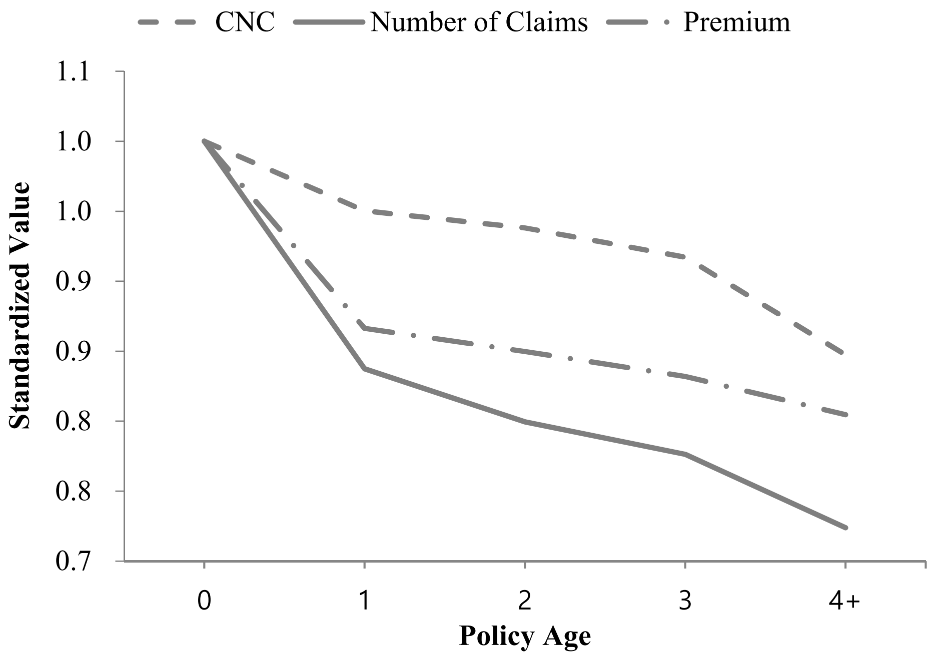
With Imperfect Information, Ph.D. Thesis, Vrije Universiteit Amsterdam, World Wide Web:
http://dare.ubvu.vu.nl/bitstream/handle/1871/13113/8631.pdf.

Zavadil, Tibor, 2015, Do the Better Insured Cause More Damage? Testing for Asymmetric Information in Car Insurance, *Journal of Risk and Insurance*, 82(4): 865–889.

Roh, Sae-Jin, and Ha Tae-Wook, 2008, An Empirical Analysis on Adverse Selection in the Korean Automobile Insurance Market, *Korean Journal of Business Administration*, 13: 41-64.

Kim, Yong-Duk, and Kyoung-Hwan Kim, 2009, A Dynamic Empirical Analysis of Adverse Selection in the Automobile Insurance Market, *Korean Insurance Journal*, 82: 279-305.

FIGURE 1
Coverage Choice, Claim Frequency, and Premium by Policy Age



Note: Figure 1 shows how the value of CNC, claim frequency, and premium by policy age. CNC is a dummy which is equal to 1 if a policyholder chooses Collision and Comprehensive coverage. All of the values of the three variables are standardized by the value of Group0, of which policy age is less than one year.

TABLE 1
Standards for Surcharging Premiums

Type of Claims			Points per Claim	Change in BM Class
Bodily Injury	Death		4	1 class down for each point Maximum of 6 classes down
	Injury	Degree 1	4	
		Degree 2 to 7	3	
		Degree 8 to 12	2	
		Degree 13 to 14	1	
Non-Liability Claims			1	
Property Damage	Payment > Threshold		1	
	Payment < Threshold		0.5	No surcharge for initial claim

Note: Table 1 shows how to surcharge a policyholder's premium depending on his or her claim history.

TABLE 2
Descriptive Statistics

Variable	Mean: All	Policies with Collision and Comprehensive	Policies without Collision and Comprehensive
Policyholder characteristics			
Age	44.813	44.220	46.511
	(11.287)	(11.180)	(11.417)
Gender	0.751	0.743	0.774
	(0.432)	(0.437)	(0.418)
Bonus-Malus coefficient	16.043	16.004	16.156
	(4.573)	(4.519)	(4.710)
Driving experience	5.647	5.555	5.908
	(2.278)	(2.348)	(2.041)
Policy age	1.577	1.499	1.798
	(1.445)	(1.420)	(1.490)
Policyholder's car characteristics			
Car age	6.218	5.004	9.687
	(4.415)	(3.870)	(4.018)
Type	2.001	2.016	1.961
	(1.248)	(1.245)	(1.255)
Value of the car	11,067,057	13,094,036	5,273,279
	(11,473,341)	(12,086,561)	(6,710,500)
Domestic	0.952	0.944	0.974
	(0.214)	(0.230)	(0.159)
Realization of policy			
Number of claims	0.149	0.150	0.145
	(0.399)	(0.400)	(0.396)
Premium	350,440	347,577	358,630
	(150,240)	(144,570)	(165,116)
N	584,778	433,301	151,477

Note: Table 2 contains important characteristics of both policyholders and their cars, as well as policy performance.

TABLE 3
Definition of Variables

Dependent variables	
CNC	1 if Collision and Comprehensive coverage is included in the policy
Claim occurrence	1 if there is at least one claim for Bodily Injury I and Property Damage
Independent variables	
Age1 to Age7	age of the primary driver of the insured car, categorized into 7 dummies
Age limit1	1 if no special contract on age is included
Age limit2	1 if special contract on age over 21 is included
Age limit3	1 if special contract on age over 24 is included
Gender	1 if primary driver is male
BM1 to BM4	coefficient from Bonus-Malus Scheme, categorized into 4 dummies
EX0 to EX5	driving experience (years since starting to drive), categorized into 6 dummies
Policy age0 to Policy age3	the length of time under contract, categorized into 4 dummies
Car age0 to Car age9	years the car has been used, categorized into 10 dummies
Type1 to Type4	type of car, categorized into 4 dummies
Value	log-transformed value of the insured car, in Korean won
Domestic	dummy variable, 1 if the car is domestic
Premium	log-transformed value total premium for Bodily Injury I and Property Damage, in Korean won

Note: Table 3 contains all the variables used in the probit models. Independent variables are included for conditioning on correlation between two dependent variables.

TABLE 4
Result of Probit Regressions

	Dependent Variables			
	Collision and Comprehensive Coverage		Claim Occurrence	
Independent Variable	Coef.	Std. Err.	Coef.	Std. Err
Age1	-0.399**	0.154	-0.372**	0.156
Age2	-0.043	0.034	-0.137***	0.035
Age3	0.163***	0.032	-0.207***	0.034
Age4	0.255***	0.030	-0.248***	0.033
Age5	0.195***	0.030	-0.160***	0.033
Age6	0.178***	0.031	-0.149***	0.033
Age7	0.110***	0.032	-0.098***	0.034
Age limit1	-0.187***	0.047	0.518***	0.043
Age limit2	-0.246***	0.018	0.335***	0.017
Age limit3	-0.126***	0.015	0.193***	0.014
Gender	-0.105***	0.005	-0.005***	0.005
BM1	-0.811***	0.029	0.579***	0.029
BM2	-0.358***	0.009	0.401***	0.010
BM3	-0.239***	0.006	0.239***	0.007
BM4	-0.047***	0.006	0.110***	0.006
EX0	-0.060***	0.012	0.220***	0.011
EX1	-0.050***	0.011	0.063***	0.011
EX2	-0.003	0.010	-0.001	0.010
EX3	-0.006	0.011	-0.016	0.011
EX4	-0.024**	0.011	-0.029***	0.011
EX5	-0.050***	0.010	-0.037***	0.011
Policy age0	-0.044***	0.006	0.071***	0.007
Policy age1	-0.119***	0.006	0.051***	0.007
Policy age2	-0.088***	0.007	0.041***	0.008
Policy age3	-0.094***	0.008	0.035***	0.009
Car age0	1.232***	0.017	-0.136***	0.016
Car age1	1.025***	0.015	-0.165***	0.015
Car age2	0.861***	0.013	-0.133***	0.014
Car age3	0.752***	0.012	-0.124***	0.014
Car age4	0.599***	0.012	-0.095***	0.013
Car age5	0.577***	0.011	-0.047***	0.012
Car age6	0.526***	0.010	-0.048***	0.012
Car age7	0.442***	0.010	-0.028**	0.011
Car age8	0.377***	0.009	-0.015	0.011
Car age9	0.248***	0.008	0.007	0.010
Type1	0.000***	0.009	-0.134***	0.009
Type2	0.030***	0.007	-0.089***	0.007

Type3	-0.103***	0.006	-0.041***	0.006
Type4	-0.105***	0.008	-0.032***	0.008
Value of the car	0.403***	0.005	0.051***	0.005
Domestic	0.329***	0.011	0.012***	0.011
Wald test of $H_0 = 0$	20.665	0.000		
ρ of bivariate probit	0.028	0.006		

Note: Table 4 presents the estimates of coefficients and standard errors from two probit models: one for the choice of Comprehensive and Collision coverage (a dummy equal to 1 if a policyholder chooses such coverage) and the other for claim occurrence (a dummy equal to 1 if there is at least one claim for Bodily Injury I and Property Damage). Among results from the bivariate probit model, Table 4 only contains the estimate of correlation coefficient ρ, because all other estimates are very similar to those from the first test. As was shown in Table 3, independent variables consist of the characteristics of policyholders and of their cars. By using these variables, I investigate evidence for conditional correlation between two dependent variables. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

TABLE 5
Coverage Choice and Number of Claims by Driving Experience

	Coefficient for Collision and Comprehensive								
Model	All	0	1	2	3	4	5	6	7+
OLS									
(1)	0.008***	0.063***	0.015**	0.010	0.005	0.008	0.007	0.006	0.007***
	(0.001)	(0.010)	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.001)
(2)	0.009***	0.031***	0.004	-0.002	-0.001	0.003	0.006	0.004	0.007***
	(0.001)	(0.008)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.001)
Poisson									
(3)	0.059***	0.234***	0.077*	0.056	0.035	0.053	0.045	0.041	0.050***
	(0.009)	(0.036)	(0.040)	(0.038)	(0.043)	(0.043)	(0.044)	(0.045)	(0.011)
(4)	0.064***	0.103***	0.021	-0.012	-0.007	0.019	0.055	0.037	0.053***
	(0.008)	(0.031)	(0.034)	(0.033)	(0.037)	(0.037)	(0.038)	(0.040)	(0.010)
N	584,778	32,482	27,746	31,534	24,582	23,863	23,982	21,963	398,626

Note: Table 5 presents the estimates of coefficients and standard errors from OLS and Poisson regressions. The dependent variable is the number of claims for Bodily Injury I and Property Damage. The independent variable indicates the choice of Comprehensive and Collision coverage (a dummy equal to 1 if a policyholder chooses such coverage). Models (1) and (3) include the set of independent variables shown in Table 3 as control variables. Models (2) and (4) only consider premium as a control variable. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

TABLE 6
Coverage Choice and Number of Claims by Policy Age

	Coefficient for Collision and Comprehensive					
Model	All	0	1	2	3	4+
OLS						
(1)	0.008*** (0.001)	0.014*** (0.003)	0.008*** (0.003)	0.008*** (0.003)	0.014*** (0.004)	0.002 (0.003)
(2)	0.009*** (0.001)	0.011*** (0.003)	0.006** (0.002)	0.002 (0.003)	0.012*** (0.004)	0.005** (0.002)
Poisson						
(3)	0.058*** (0.009)	0.086*** (0.016)	0.056*** (0.017)	0.051** (0.022)	0.105*** (0.031)	0.016 (0.020)
(4)	0.064*** (0.008)	0.065*** (0.014)	0.041*** (0.015)	0.018 (0.020)	0.089*** (0.027)	0.050*** (0.019)
N	584,778	173,573	161,167	94,564	50,474	10,500

Note: Table 6 presents the estimates of coefficients and standard errors from OLS and Poisson regressions. The dependent variable is the number of claims for Bodily Injury I and Property Damage. The independent variable indicates the choice of Comprehensive and Collision coverage (a dummy equal to 1 if a policyholder chooses such coverage). Models (1) and (3) include the set of independent variables shown in Table 3 as control variables. Models (2) and (4) only consider premium as a control variable. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.