Does Less Risk Classification Induce More Adverse Selection?: Evidence from Automobile Insurance Market

KUNIYOSHI SAITO*

March, 2004

Abstract

Recent empirical studies in the competitive automobile insurance markets show there are no signs of adverse selection in these markets. This paper examines whether adverse selection could be induced by the rate regulation which prohibits insurance companies from using some of the drivers' attributes. Using an individual data set from a heavily regulated automobile insurance market, we obtained several findings as follows. Firstly, no evidence of adverse selection was found in general: positive correlation between risk and coverage was not found for either beginners or experienced drivers. Secondly, this result is robust in the sense that it holds under several empirical procedures and different definitions of risk and coverage. Thirdly, we find that risk-related variables do not necessarily induce adverse selection: the null hypothesis that consumers in risky regions are more likely to purchase insurance was tested against its alternative and rejected. Our study supports the recently prevailing view that adverse selection phenomenon exists only to a very limited extent at least in this market.

Key words: adverse selection, risk classification, rate regulation, automobile insurance

JEL Classification: D81, D82, G22, L51

*Graduate School of Economics, University of Tokyo. 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Phone, Fax: (81)-3-5841-5588. E-mail: ksaito@grad.e.u-tokyo.ac.jp. The earlier version of this paper corresponds to the old manuscript entitled "Does Less Risk Classification Induce more Adverse Selection?: Testing for Asymmetric Information in Auto Insurance Market." I have benefited from the comments of the seminar participants at the University of Tokyo and the annual meeting of Japanese Economic Association held at Meiji University. I am also indebted to an insurance company and its employee who made this data set available. Any errors are my own responsibility.

1 Introduction

Recent empirical studies in the competitive automobile insurance markets show there are no signs of adverse selection in these markets. To take some of the outstanding studies, Chiappori and Salanié (2000a) find no systematic relationship between risk and coverage in French automobile insurance market and Dionne *et al.* (2001) propose much the same results in Quebec. Although Cohen (2002) finds a strong relationship between risk and coverage using a rich data set from Israelite market, it holds only for the groups of drivers whose driving experiences are relatively longer than the insurance company.[1] One straightforward interpretation of these results would be that, whatever the reasons are, insurers are successfully managing the problems caused by asymmetric information, at least in the competitive insurance markets where insurers can freely offer the menu of contracts.

Insurance premiums, however, are often regulated for various economical and political reasons. Two types of regulation seem to be common in automobile insurance markets. One is the 'rate compression', which prohibits insurance companies from using some of the attributes of the policyholder, such as sex, region, or marital status *etc.* The other is the 'rate suppression', which restricts the maximum level of premiums applied to particular risk categories.[2]

In many cases, these regulations have succeeded in providing more coverage for the high risk agent who would otherwise be offered an extremely high premium and results in being little or no coveraged. On the other hand, these types of regulation can cause a serious adverse selection problem because insurers cannot offer the menu of contracts which reflect each agent's risk level under such a constraint.[3]

In spite of the wide prevalence of rate regulation in automobile insurance market, there has been few empirical studies which intends to evaluate the impacts of rate regulation on the adverse selection or moral hazard. One

[1]It should be also noted that her data comes from an entrant. See Chiappori and Salanié (2000b), Löfgren *et al.* (2002), and Siegelman (2003) for broader surveys.

[2]See Harrington (1992) for more discussions about this regulation.

[3]For the arguments of risk classification in insurance markets, see Bond and Crocker (1991), Crocker and Snow (1986, 2000), Dahlby (1983), Hoy (1982), Jaffee and Russell (2002), Puelz and Kemmsies (1993), and Rea (1992)

exception is Dahlby (1983), who examines the effects of rate prohibition on sex and shows that it induces adverse selection using the data from Canada. Our study differs from and improves on his study at least two aspects. One is that we use individual data, while Dahlby uses aggregated data. The other is that we use the actual claim data, while Dahlby's results are based on the simulation method.

The other characteristics of this study can be summarized as following two aspects. Firstly, we test adverse selection in various aspects. Although most of the previous studies have focused on the deductibles which are at most $1,000 or so, we use several definitions of coverage and risk to check the robustness of our basic result. Even if we do not find any evidence of adverse selection when we define risk as occurrence of a crash accident, it can still be possible that adverse selection matters in the case of theft.

Secondly, we examine the effects of rate prohibition *directly*. If there is no risk classification by region and if the adverse selection story holds, consumers in high risk regions should be more likely to purchase insurance contracts than those in less accident regions. We examine whether this hypothesis is supported by our data.

Main findings can be summarized as follows. Firstly, we find no evidence of adverse selection in general: conditional on all the variables observed by insurer, the null of independence between risk and coverage is not rejected. Secondly, the result is robust in the sense that it holds under the several empirical procedures and definitions of risk including the occurrence of crash accident, theft, and amount of loss. Thirdly, our study shows that the seemingly risk-related variables do not necessarily induce adverse selection: the hypothesis that the consumers in risky regions are more likely to purchase insurance is tested against its alternative and rejected.

The remainder of this paper is organized as follows. In the next section, a simple description of Japanese auto insurance market and our empirical methods are proposed. The introduction of our data set and basic results are presented in Section 3. We check the robustness of the basic results in Section 4. Finally, concluding remarks are presented in Section 5.

2 Empirical Procedure

2.1 Automobile Insurance Market in Japan

In this section, we describe the auto insurance market in Japan which gives an ideal situation for our study.

Until the partial rate liberalization in 1998, there had been a non-profit organization called 'Automobile Insurance Rating Organization' (AIRO) in this market.[4] The role of AIRO was to gather the data from its member companies and calculate both the compulsory and voluntary automobile insurance premiums.[5] The AIRO rates had been checked by the Commissioner of the Financial Services Agency and it claimed that the rates should be 'not unfairly discriminatory'. As a result, premiums had not been discriminated by some of the attributes of the driver such as region, sex, profession, or annual mileages, which all seem to have strong correlation with drivers' risk. The maximum level of premium had been also suppressed. Since most of the insurance companies joined AIRO and it was forced for its members to use the AIRO rates by law, consumers had no choice but to purchase the contracts calculated by AIRO. From these aspects, Japanese automobile insurance market provides an ideal situation for our purpose.

2.2 Empirical Procedure and Variable Definitions

As an empirical procedure, we basically follow Chiappori and Salanié (2000a). They argue that when coverage and risk are defined as y_i and z_i, respectively, the test of adverse selection can be simply expressed as a positive correlation between y_i and z_i.[6] Since the correlation between y_i and z_i only holds within

[4] For more closer information about the Japanese automobile insurance market, see AIRO (2002), Dionne (2002), and Hayakawa *et al.* (2002).

[5] Although joining to the AIRO was not compulsory, most of the insurers became members.

[6] Chiappori *et al.* (2002) calls this property 'positive correlation property.' Since they argue this property basically holds under the competitive market, (or more precisely, the zero-profit condition), and since the automobile insurance market in Japan is far from competitive as we describe in the body, we have to confirm the appropriateness of applying this property to Japanese auto insurance market. As they argue, the zero-profit condition can be weakened to the condition that contracts with broader coverage have (weakly) less profit. We checked whether this condition holds in this market and found the premium-loss ratio is consistently higher for broader coveraged contracts, which suggests that broader coverage

the group of drivers insurance companies regard as the 'same', the correlation should be tested conditional on all the variables insurance company observes about their policyholders. We denote this set of observed variables as X_i following Chiappori and Salanié (2000a).

Throughout this paper, we concentrate on two types of voluntary insurance contracts: collision insurance and its deductibles. This is because the choice of these coverages are especially important in the market under consideration, and because there has been an anecdotal evidence that adverse selection is more likely to matter in the case of first-party insurance.[7] Therefore, the coverage of insurance contract y_i is defined as following two ways.

$$
\begin{aligned}
y_i^c &= \begin{cases} 1 & \text{if agent } i \text{ purchased collision insurance} \\ 0 & \text{otherwise} \end{cases} \\
y_i^d &= \begin{cases} 1 & \text{if agent } i \text{ purchased collision insurance with zero-duductible} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

where y_i^c defines the choice of collision insurance, and y_i^d defines the level of deductibles.[8]

We define the risk variable z_i as follows.

$$
z_i^c = \begin{cases} 1 & \text{if agent } i \text{ had at least one crash accident} \\ 0 & \text{otherwise} \end{cases}
$$

where z_i^c is defined as whether policyholder i had at least one crash accident. We define risk in this way because of the following 'accident $\neq$ claim' problem. Let me consider the case in which agent i's car was stolen. If he had a collision

have less profit.

[7]Two package contracts are common in Japanese automobile insurance market. One is a narrow coverage contract called PAP, which includes the coverages for bodily injury liability, property damage liability, and passengers' personal accident. The other one is a broad coverage contract called SAP, which includes collision insurance in addition to the coverages in PAP. Whether or not to purchase voluntary insurance seems not an important decision making, because most drivers purchase either one of these two package contracts. However, which one of these to choose seems significant, because only less than 50 percent private-use cars drivers purchase SAP contracts. Also, PAP and SAP largely differ in their premiums. In 1997, the average premium for SAP was 85,750 yen and that for PAP was 51,247 yen.

[8]More explanation about deductibles would be helpful. There are two cases which correspond to $y_i^d = 1$. One is the case that policyholders purchased the contract with 0-10 proportional contract. The other is the case that policyholders purchased the contract with zero-duductible options which is only available for those who purchased 5-10 proportional contracts and belong to above 7-class experience ratings. We defined $y_i^d = 1$ if agent i purchased one of these contracts.

insurance, he could make a claim for insurance company, but if he had no collision insurance, no claim will be listed in the insurer's file because it is not covered by the contract originally. If we do not control this bias, we will mistakenly conclude that positive correlation was found between coverage and risk. To avoid this 'accident $\neq$ claim' problem, we have to define the risk which will always become a claim whatever type of contract he has. In the above equation, this problem is mitigated since every policy covers the crash accident.[9] For the same reason, we set $z_i^c = 0$ for the accident with a loss of less than 50 thousand-yen.

Finally, four variables are included in X_i: type of cars (3) (private-use, standard size cars, *etc.*), age of main drivers (4) (not covered less than 21 year old drivers, *etc.*), class of collision insurance (10) (class1-class 9 or no-class, according to the type or age of the car), and size of car (5) (less than 1500cc engine size *etc.*) (() denotes the number of classifications).[10]

Under these definitions of risk z_i, coverage y_i, and observations by insurer X_i, we test the conditional independence between y_i and z_i.

3 Data and Preliminary Results

3.1 The Data

Our data is from one of the large insurance companies in Japan. It includes 30,000 voluntary insurance policies undertaken during the period from April 1st, 1999 to March 31st, 2000.[11]

The data was extracted using stratified sampling. As the first step, we categorized all the prefectures using four criteria: 1. diffusion rate of collision insurance (1999), 2. claim rate of collision insurance (1999), 3. population density (1999), 4. eastern or western area in Japan. From these stratum,

[9] More precisely, every policy has the coverages for property damage liability insurance. Since more than 90 percent of claims for property damage liability insurance are caused by crash accidents, we describe z_i as the occurrence of at least one crash accident.

[10] Family restriction discount and safety devices discount also affect the premium. These variables are not included because the data was not available.

[11] Since the rate liberalization was in 1998, it was already allowed for insurers to use some additional variables. However, the fee structure was as much the same as that in the pre-deregulation years at the time of this sample period.

Table 1: Characteristics of each Prefecture

Prefecture	Diffusion Rate of Collision Insurance	Claim Rate of Collision Insurance	Population Density	Eastern or Western Area
Iwate	24.7%	7.04%	93	East
Kanagawa	38.0%	8.66%	3,448	East
Osaka	42.8%	9.52%	4,650	West
Kagoshima	22.9%	5.67%	196	West

Note: All the figures are in 1999 values.
Source: Automobile Insurance Rating Organization (1999) ('Diffusion rate' and 'Claim rate') and Statistics Bureau in Japan (1999) ('Population Density')

four representative prefectures were selected: Iwate, Kanagawa, Osaka, and Kagoshima.[12] The characteristics of each prefecture are summarized in Table 1. As the second step, using the prefecture-level diffusion rate and average claim rate of collision insurance, we calculated the number of contracts which is expected to include sufficient number of accidents. With this total number of contracts fixed, the data was extracted from each prefecture proportionally to the number of voluntary insurance contracts. These two steps give the data: Kanagawa (12,246), Osaka (12,829), Iwate (2,152), Kagoshima (2,773), Total (30,000).[13]

All the policies have following fifteen variables: the type of contract, the age of main drivers, class of experience ratings, type of car, size of car, maximum coverage of bodily injury liability insurance, maximum coverage of property damage insurance, level of deductibles for property damage insurance, maximum coverage for automobile passengers' personal accident insurance, with or without the collision insurance, optional choice for the collision insurance, level of deductibles for the first and second claim (for collision insurance), with or

[12]We extracted the data from four prefectures only because it was too costly to extract the data from whole country.

[13]Three caveats would be worth mentioning for the data. Firstly, fleet contracts are excluded. Hence, all the policies are contracted with less than ten vehicles in one contract, which means we only focus on individual purchasers or comparatively small firms. Secondly, coinsurance contracts are excluded. That is, contracts undertaken by several insurers are excluded and we limit to the sample undertaken by one insurer. Finally, in our data, there are little 'incurred but not reported' (IBNR) accidents. It sometimes takes several years for an accident to be settled in the data file, especially in the case of bodily injury accident. Since the policies were contracted in 1999 and the data was extracted in February 2003, almost all the claims are fixed at the time of data extraction.

without zero-deductible options, class of collision insurance, and age of cars. In addition to these variables, contracts with claims include the following five variables: number of accidents, type of insurance claimed, the person who had accidents, type of accident, and amount of loss paid. Thus, our data includes the detailed information about the contracts (y_i), accidents (z_i), and observation by insurer (X_i), which are all necessary for the test of adverse selection and/or moral hazard.

Since the data set includes several types of cars (commercial-use cars, agricultural-use cars, motor cycles *etc.*), we limit our sample to private-use cars: there are 21,997 such contracts. In addition, we divide the data into 'beginners' and 'experienced drivers' because of the following two reasons. Firstly, the accuracy of risk perception could be different according to the driving experience. Drivers might have more accurate knowledge about their accident probability as they have more driving experience. Secondly, drivers with more experience could be better knowledged about the contracts they purchase. Policyholders may not be able to evaluate the contents of the contract for their first purchases. In such cases, consumers' action could vary according to their driving experience.

We use the experience rating class as a proxy for the drivers' experience.[14] A policyholder belongs to 6-class for the first year, and 3 classes down if he had a claim, and 1 class up if they had no claim. This system suggests that most of the policyholders in 6- or 7-class are those who have less than 2 years driving experience.[15] Hereafter, we call the policyholders who belong to 6- or 7-class as 'beginners', and those who belong to all the other classes as 'experienced drivers'. Under these definitions, we have 2,813 contracts for beginners, and 19,184 contracts for experienced drivers.[16]

[14]This is only because our data set does not include the variables which directly indicate the drivers' experience such as the year of obtaining the driver's license.

[15]Needless to say, there could be some drivers who once belonged to other classes and then came back to 6- or 7-class. However, such cases are not so many in our data according to the insurer.

[16]We do not use the driver's age as a proxy for experience, only because it is not included in the data.

Table 2: Cross Tabulations

Collision Insurance

Total

y_i^c	z_i^c = 0	1	Total	Claim Rate
0	9,972	519	10,491	4.95%
	47.80	45.81	47.69	
1	10,892	614	11,506	5.34%
	52.20	54.19	52.31	
Total	20,864	1,133	21,997	5.15%
	100	100	100	

$\chi^2 = 1.702$ $Pr = 0.192$

Deductibles

Total

y_i^d	z_i^c = 0	1	Total	Claim Rate
0	1,443	107	1,550	6.90%
	13.25	17.43	13.47	
1	9,449	507	9,956	5.09%
	86.75	82.57	86.53	
Total	10,892	614	11,506	5.34%
	100	100	100	

$\chi^2 = 8.706$ $Pr = 0.003$

Collision Insurance

Beginners

y_i^c	z_i^c = 0	1	Total	Claim Rate
0	1,600	112	1,712	6.54%
	60.98	59.26	60.86	
1	1,024	77	1,101	6.99%
	39.02	40.74	39.14	
Total	2,624	189	2,813	6.72%
	100	100	100	

$\chi^2 = 0.218$ $Pr = 0.641$

Deductibles

Beginners

y_i^d	z_i^c = 0	1	Total	Claim Rate
0	414	36	450	8.00%
	40.43	46.75	40.87	
1	610	41	651	6.30%
	59.57	53.25	59.13	
Total	1,024	77	1,101	6.99%
	100	100	100	

$\chi^2 = 1.185$ $Pr = 0.276$

Collision Insurance

Experienced Drivers

y_i^c	z_i^c = 0	1	Total	Claim Rate
0	8,372	407	8,799	4.63%
	45.90	43.11	45.76	
1	9,868	537	10,405	5.16%
	54.10	56.89	54.24	
Total	18,240	944	19,184	4.92%
	100	100	100	

$\chi^2 = 2.804$ $Pr = 0.094$

Deductibles

Experienced Drivers

y_i^d	z_i^c = 0	1	Total	Claim Rate
0	1,029	71	1,100	6.45%
	10.43	13.22	10.57	
1	8,839	466	9,305	5.01%
	89.57	86.78	89.43	
Total	9,868	537	10,405	5.16%
	100	100	100	

$\chi^2 = 4.205$ $Pr = 0.040$

3.2 A First Look at the Data

To begin with, we overview our data set. The findings in this subsection should be only viewed as a descriptive statistics.

Table 2 shows the cross tabulations for collision insurance and its deductibles. Three tables in the left column of Table 2 report the cross tabulation for collision insurance. Out of 21,997 policies, 52.31 percent purchased collision insurance. The claim rate is slightly higher for those with collision insurance: 5.34 percent for $y_i^c = 1$ group and 4.95 percent for $y_i^c = 0$ group, which does not imply the existence of adverse selection.

Much the same results are found for beginners and experienced drivers. For beginners, the average claim rate is 6.99 percent (= 77/1,101) and 6.54 percent (= 112/1,712) for those with and without the collision insurance, respectively. In the case of experienced drivers, the correspondence average claim frequencies are 5.16 percent (= 537/10,405) and 4.63 percent (= 407/8,799). In each group, however, the χ^2 test statistics are small and the null of independence between y_i and z_i is rejected at a 10 percent statistical level only for experienced drivers.

Three tables in the right column of Table 2 show that the tendency is slightly different for the deductibles. Out of 11,506 policies with collision insurance, 86.53 percent drivers purchased collision insurance with zero-duductible contract. The average claim rate is 5.09 percent (= 507/9,956) for the group with zero-duductible, and 6.90 percent (107/1,550) for the group with positive deductibles. The claim rate is higher for those with positive deductibles and the difference is statistically significant at a 1 percent level ($\chi^2 = 8.706$). We find much the same tendencies for beginners and experienced drivers, although the difference is not statistically significant for beginners. This suggests that the drivers with more coverage are less likely to have an accident, which is the opposite sign of adverse selection story.[17]

To summarize, a first look at our data set shows that: (1) we find weak positive correlation between y_i and z_i for collision insurance, which is not

[17] This result seems to support the story of 'propitious selection' proposed by Hemenway (1990, 1992): when risk averse drivers tend to purchase more insurance and drive more cautiously, the correlation between risk and coverage becomes negative. However, we leave the discussion of this story to future research.

statistically significant, (2) we find negative correlation between y_i and z_i for deductibles, which is statistically significant only for experienced drivers.

3.3 Basic Results

Simple examination of our data so far suggests that the relationship between y_i and z_i is positive for collision insurance and negative for deductibles although the former relationship is not statistically significant. In order to test adverse selection, however, we have to control for all the variables observed by the insurer since the independence of risk and coverage should be examined *within* the group of drivers which the insurer regards as the same risk class. To this end, two empirical methods are used to test the conditional independence between y_i and z_i: bivariate probit model and χ^2 test, both of which are proposed and recommended by Chiappori and Salanié (2000a).

A. Bivariate Probit

We define the usual bivariate probit model as follows,

$$\begin{aligned} y_i &= \begin{cases} 1 & \text{if } y_i^* = \beta X_i + \epsilon_i > 0 \\ 0 & \text{otherwise} \end{cases} \\ z_i &= \begin{cases} 1 & \text{if } z_i^* = \gamma X_i + \eta_i > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where the first equation describes the choice of contract, and the second equation describes the occurrence of an accident. Two error terms ϵ_i and η_i satisfy the following standard conditions: $E[\epsilon_i] = E[\eta_i] = 0$, $Var[\epsilon_i] = Var[\eta_i] = 1$, and $Cov[\epsilon_i, \eta_i] = \rho$. We estimate these models for collision insurance (y_i^c) and deductibles (y_i^d).

Table 3 shows the estimation results. Here, only the estimated value of ρ are reported since our interest is in whether or not the value of ρ is statistically different from zero. The table suggests that the results of simple analyses in the previous section are not altered even if we control for the set of variables X_i. The first two columns of Table 3 provide the results for collision insurance. Although the signs of ρ are positive in each group, the estimated values of ρ are all close to zero and do not reject the null of conditional independence between y_i and z_i (0.062 for beginners and 0.031 for experienced drivers).

Table 3: Bivariate Probit Estimation Results

	Collision Insurance		Deductibles	
	Beginners	Experienced	Beginners	Experienced
ρ	0.062	0.031	-0.076	-0.065*
s.e.	0.048	0.020	0.075	0.034
95 % conf. int.	[-0.032, 0.155]	[-0.008, 0.069]	[0.219, 0.072]	[-0.131, -0.001]
N	2,813	19,184	1,101	10,405

Note: * indicates significant at a 10% level.

Columns 3 and 4 give the results for deductibles. The signs of ρ are negative for both beginners and experienced drivers, and the null of independence is rejected at a 10 percent level for experienced drivers. It suggests that the low risk drivers are likely to choose broader coverage which is an opposite implication from adverse selection. Therefore, the bivariate probit models show that no evidence of adverse selection is found either for collision insurance or its deductibles.

B. χ^2 Test

Bivariate probit model is valid under some specific assumptions such as linearity of the latent variable equations and normality of the error terms. It has been pointed out, however, that the model often gives wrong results when these conditions are not satisfied. In order to check the robustness of our result so far, we test the conditional independence of y_i and z_i by χ^2, following Chiappori and Salanié (2000a).

Firstly, we divide our data into several groups using variables in X_i. In particular, four variables are used for the categorization: age of main driver, type of car, class of car, and size of car. Using the criteria shown in Table 4, the data was divided into $2^4 = 16$ cells.[18] In each of the cell, we construct cross tabulations and calculate the χ^2 statistics (Q), which follows a $\chi^2(1)$ distribution. We also calculate the statistics S, defined as the summation

[18]The criteria in Table 4 was determined so that each cell has the sufficient number of samples in each cell. However, we could not avoid to have too few policies to calculate the χ^2 statistics.

Table 4: Definitions of variables in X_i

Variables	Definition
AGE = 0	Covered more than 30
AGE = 1	otherwise
TYPE = 0	Small Car
TYPE = 1	otherwise
CLASS = 0	4-Class or Below
CLASS = 1	otherwise
SIZE = 0	1500cc-2500cc
SIZE = 1	otherwise

of Q, which also follows the χ^2 distribution under the null of independence between y_i and z_i.[19]

Table 5 shows the estimated values of Q in each cell. The left hand side of the table shows that the values of Q are close to zero in most of the cells. It is statistically significant only in one cell: (1,0,0,0) of the experienced drivers.[20] When we further categorize this cell using additional variables in X_i, however, we find no more correlation. Therefore, we conclude that positive correlation is found in no cell. The S statistics shows the null of independence between y_i and z_i is not rejected for the overall sample, either.

The right hand side of Table 5 shows the results for deductibles. It shows that the values of Q are also small for deductibles. We find statistically significant Q values for three cells: (1,1,0,0) for beginners, (0,0,0,1) and (1,1,1,1) for experienced drivers. Although these correlation did not disappear even when we include additional variables in X_i, it should be noted that the correlations between y_i and z_i are *negative*, suggesting that the drivers with zero-duductible are less likely to have an accident. The S statistics, however, suggest that there is no adverse selection as a whole.

To summarize, the χ^2 test procedures show that, (1) we find no evidence of adverse selection for either collision insurance or deductibles in general, because

[19] Since we do not have $y_i = 1$ or $z_i = 1$ observations for some groups, such groups are excluded in calculating the degree of freedom of S statistics. Therefore, the degree of freedom varies according to the components of the summation.

[20] Some information about this group would be helpful. The number of policies with collision insurance is 986 out of 2,116 total sample. The claim rate of $y_i = 1$ group is 6.39 percent (=63/986) and that of $y_i = 0$ group is 4.61 percent (=47/1020). The correlation comes from the group of, small-car, not covering less than 21 year-old drivers.

all the S statistics are insignificant, (2) for deductibles, we find statistically significant negative correlation between y_i and z_i for some cells. These results again show that we find no evidence of adverse selection by the χ^2 test, either.

Table 5: χ^2 Test Results

	Collision Insurance				Deductibles			
	Beginners		Experienced		Beginners		Experienced	
Group	Q	N	Q	N	Q	N	Q	N
(0,0,0,0)	0.962	202	0.005	2,567	0.939	74	0.554	1,478
(1,0,0,0)	0.033	335	2.751*	2,116	0.160	113	0.207	1,049
(0,1,0,0)	0.82	64	0.159	1,016	1.126	34	1.862	682
(0,0,1,0)	1.714	75	0.226	839	0.157	25	0.021	410
(0,0,0,1)	0.178	203	1.836	2,047	1.548	80	3.262*	1,137
(1,1,0,0)	0.192	125	0.006	899	3.758*	62	0.662	548
(1,0,1,0)	0.103	241	0.601	903	0.013	59	0.691	334
(1,0,0,1)	0.268	329	0.065	1,992	0.054	130	2.349	961
(0,1,1,0)	0.016	66	0.392	709	1.193	35	0.014	493
(0,1,0,1)	—	20	0.007	291	—	10	0.495	173
(0,0,1,1)	—	8	0.635	54	—	0	—	9
(1,1,1,0)	0.455	88	0.059	574	2.002	41	1.305	311
(1,1,0,1)	0.599	30	0.313	257	0.950	19	2.079	147
(1,0,1,1)	0.865	18	0.36	41	—	5	—	6
(0,1,1,1)	1.975	389	2.626	2,425	0.313	165	0.577	1,394
(1,1,1,1)	1.815	620	0.315	2,454	1.831	249	2.949*	1,273
S	9.995	2,813	10.358	19,184	14.043	1,101	17.027	10,405

Note: * indicates significant at a 10% level. Since there is no $y_i = 1$ or $z_i = 1$ cases for some cells, we can not calculate Q statistics. Such cases are denoted by '—'.

4 Different Definitions of Risk

So far, we have defined risk as the probability of having at least one crash accident. However, we can define risk in other ways such as the amount of loss if an accident occurred or the probability of having the car stolen. In this section, we check the robustness of our results using different definitions of risk.

4.1 Amount of Loss Paid

In this subsection, we redefine risk as the amount of loss paid for crash accident and examine whether there is a significant difference between the agents with different coverages.

Table 6 reports the average loss paid for beginners and experienced drivers. In this table, all the accidents with less than 50 thousand yen for policyholders with zero-deductible contracts are discarded (5 cases). Also, we added 50 thousand yen for loss paid for those without zero-deductible contracts. These adjustments are due to the 'accident $\neq$ claim' problem. In addition, we discarded the contracts with loss more than 10 million yen loss as outliers (3 cases) and the second or third claims were ignored.

In the case of collision insurance, the average loss amount is higher for $y_i = 1$ group only for experienced drivers, but the difference is small. As for deductibles, the average loss amount is higher for $y_i = 0$ group, which is contrary to the hypothesis that drivers who are likely to have large loss amount is likely to purchase more coverage. To examine whether these differences are statistically significant or not, two empirical procedures are conducted: two sample test and Wilcoxon's rank sum test.

Test results are shown in Table 7. Two sample test suggests the mean difference is only statistically significant at a 5 percent level for one test: deductibles of experienced drivers. Note, again, that the relationship is *negative*, which contradicts to the adverse selection hypothesis. Wilcoxon's rank-sum test, however, does not support this significance. The null of equal median for $y_i = 1$ and $y_i = 0$ group is not rejected in every test.

Table 6: Average Amount of Loss for Beginners and Experienced Drivers

		Collision Insurance			Deductibles		
Group	y_i^c	Num. of Claims	Mean	S.E.	Num. of Claims	Mean	S.E.
Beginners	$y_i^c = 0$	112	927,618	177,058	36	456,155	78,434
	$y_i^c = 1$	77	627,710	55,516	41	453,058	61,962
Experienced	$y_i^c = 0$	407	703,784	41,525	71	817,388	189,478
	$y_i^c = 1$	537	732,172	57,803	465	528,181	33,535

Note: Mean values are calculated using the amount of loss paid for collision insurance.

Table 7: Two-sample and Rank-Sum Test Results

		Collision Insurance		Deductibles	
Group	y_i^c	Two-Sample Test	Rank-sum Test	Two-Sample Test	Rank-sum Test
Beginners	$y_i^c = 0$	1.616	0.395	0.031	−0.358
	$y_i^c = 1$	0.108	0.693	0.975	0.721
Experienced	$y_i^c = 0$	−0.399	1.117	2.556**	0.437
	$y_i^c = 1$	0.690	0.264	0.011	0.662

Note: ** indicates significant at a 5% level. The null for two-sample test is equal mean for $y_i^d = 1$ and $y_i^d = 0$ group. The null for Wilcoxon's rank-sum test is equal median for these groups.

We also regressed the choice of contract y_i^c (or y_i^d) on all the attributes of the policyholders X_i and the amount of loss in logarithms $\log(z_i^{lc})$ (or $\log(z_i^{ld})$, respectively). Estimation results by Probit models show that the estimated coefficients on the loss amount ($\log(z_i^{lc})$ and $\log(z_i^{ld})$) are not statistically significant in either of the models.[21] Therefore, the regression models also support the results by two sample test and Wilcoxon's rank-sum test.

To summarize, both the parametric and non-parametric empirical procedures show that the positive correlation between the choice of deductibles and the amount of loss is found neither for collision insurance nor its deductibles.

4.2 Theft

Finally, we define risk as the probability of having his/her car stolen. Since the risk of car theft is covered by collision insurance in Japan, we redefine the coverage and risk as follows.

$$y_i^d = \begin{cases} 1 & \text{if agent } i \text{ purchased collision insurance with zero-duductible} \\ 0 & \text{otherwise} \end{cases}$$

$$z_i^s = \begin{cases} 1 & \text{if agent } i\text{'s vehicle was stolen} \\ 0 & \text{otherwise} \end{cases}$$

where y_i^d is the choice of deductibles equally defined as the previous section, and z_i^s is the risk proxy defined as whether or not the agent i's car was stolen. Although it is more reasonable to define $y_i = 1$ if agent i purchased collision insurance, the 'accident $\neq$ claim' problem makes it impossible to see the 'no collision insurance but stolen' ($y_i^d = 0, z_i^s = 1$) case. For this reason, we focus only on the deductibles.

Table 8 shows the estimation results by bivariate probit model equally modeled as before. In both groups, the estimated values of ρ are small (0.010 for beginners and -0.047 for experienced drivers), and the null of independence between y_i^d and z_i^s is not rejected. This result suggests that we do not find any evidence for adverse selection in the case of theft, either.

[21] In the case of collision insurance, the estimated coefficient on $\log(z_i^{lc})$ is -0.04 (0.12) for beginners and -0.02 (0.05) for experienced drivers. In the case of deductibles, the estimated coefficient on $\log(z_i^{ld})$ is -0.07 (0.21) for beginners and 0.08 (0.07) for experienced drivers. Standard errors are in the parentheses.

Table 8: Bivariate Probit Estimation Results (Theft)

	Beginners	Experienced
ρ	0.010	-0.047
s.e.	0.181	0.082
95 % conf. int.	[-0.331, 0.349]	[-0.205, 0.113]
N	1,101	10,405

Table 9: Number of Crash Accidents and Theft in each Prefecture

Prefecture	Crash Accidents (A)	Theft (B)	Private-use Cars (C)	(A) / (C)	(B) / (C)
Iwate	4,403	93	596,801	0.74%	0.02%
Kanagawa	52,663	3,170	2,819,094	1.87%	0.11%
Osaka	51,560	7,916	2,580,013	2.00%	0.31%
Kagoshima	9,420	232	735,784	1.28%	0.03%

Source: Number of theft is from National Police Agency (1999) and number of crash accidents and private-use cars are from ITARDA (1999).

4.3 Do Unobserved Variables Induce Adverse Selection?

When we look at the data from the police department, we can see that the probabilities of accidents and theft largely vary among four prefectures: the probabilities of crash accident and theft are higher in Osaka and Kanagawa, and lower in Iwate and Kagoshima (Table 9). If premiums are not discriminated by region and if adverse selection exists, drivers in Osaka and Kanagawa should be more likely to purchase insurance, than those in Iwate and Kagoshima.

To test this hypothesis, we estimate the model of the form:

$$y_i = \beta X_i + \sum_{j=1}^{3} \gamma_j REGION_{ij} + \epsilon_i,$$

where y_i is a dummy variable which is defined as either y_i^c or y_i^d, X_i is a set of attributes of the driver as described in section 3.3, $REGION_{ij}$ is a set of dummy variables, and ϵ_i is a normally distributed error term. In this equation, we select Iwate as a base group and include dummy variables for each of the remaining prefectures. In the presence of adverse selection, we should find positive signs for the coefficients on dummies for Osaka and Kanagawa.

Table 10 shows the estimation results by probit. Columns 1 and 2 of this table show that the estimated coefficients for $REGION_{ij}$ are not statistically significant for collision insurance. Columns 3 and 4 show that, in the case of deductibles, drivers in high risk prefectures are less likely to purchase more coverage. These results suggest that the uniform pricing across regions do not induce adverse selection which is consistent with our results so far.

5 Conclusion

In this paper, we set out to investigate whether adverse selection could be induced by the rate regulation using an individual data set from a heavily regulated insurance market. Contrary to the initial prediction, our various empirical studies show that no sign of adverse selection is found even in this environment. Rather, we find negative correlation between risk and coverage for some groups, which is the reverse of the adverse selection hypothesis.[22]

Taking the results of this study with the recent other studies in this market, it seems natural to conclude that adverse selection exists only to a very limited extent in this market as conjectured by Chiappori and Salani'e (2000a).[23] However, it should also be noted that insurers have devoted and are still now devoting large efforts to cope with the threat from adverse selection and moral hazard. Indeed, we find some evidences which suggest that insurers were suffered from adverse selection in the early stage of auto insurance market where insurers did not have so sophisticated devices as those they have now.[24] Therefore, it may be appropriate to conclude that adverse selection can cause a serious problem at least *potentially*, but insurers can mitigate such a problem by exploiting some basic mechanisms.[25]

[22] Of course, we are not insisting that there is no adverse selection or no moral hazard in this market. It is extremely easy to find an evidence which suggests the importance of informational asymmetry in this market: we have only to spread a news paper and find the articles on insurance frauds, for example.

[23] "... we do not find evidence of (risk-related) adverse selection because this phenomenon does not exist (or only to a very limited extent) in the market under consideration". (Chiappori and Salani'e (2000a), p.73)

[24] See Saito (2003) for more information about the adverse selection in the early insurance market.

[25] Also, we agree with Siegelman (2003) in that adverse selection story is often too much exaggerated.

Finally, two caveats deserve mention for our results. Firstly, although we have carefully checked the robustness of our basic result, the 'accident $\neq$ claim' problem is not overcome and the results could be largely affected by this problem. Indeed, the data from police office reports that the probability of theft is much higher for insured drivers than total average: the average percentage of theft in the four prefectures is 0.10 percent in 1999 (11,411 cars were stolen out of 11,880,389 cars), while our insurance data reports it is 0.22 percent (67 cars were stolen out of 30,000 private-use cars).[26] This implies that insured drivers are more likely to have their cars stolen.

The second caveat is that the results might be specific to the automobile insurance market. Although Cawley and Philipson (1999 propose the negative evidence of adverse selection in life insurance market, positive evidences are proposed in the credit card market (Ausubel (1999)), in the annuity market (Finkelstein and Poterba (2000)), and in the health insurance market (Barrett and Conlon (2003)). The effects of rate regulation in insurance market might also differ among different markets. For instance, Harrington and Danzon (2000) shows the rate suppression in workers' compensation insurance brings about the increase in the frequency and/or severity of employee injuries, while Buchmueller and DiNardo (2002) finds that the community rating does not induce adverse selection death spiral in health insurance market. Therefore, the test for adverse selection and the effects of rate regulation should be carefully examined in each market.

[26]Figures are not restricted on private-use cars because the number of theft from police office is only available for total car. There is no theft case for the non private-use car contracts in our data.

Table 10: Estimation Results for the Choice of Contracts (Probit Models)

	Dependent Variable:			
	Collision Insurance (y_i^c)		Deductibles (y_i^d)	
	Beginners	Experienced	Beginners	Experienced
AGE1	−0.393***	−0.308***	−0.196	−0.190***
	(0.086)	(0.040)	(0.157)	(0.076)
AGE2	0.030	−0.200***	−0.432***	−0.311***
	(0.060)	(0.026)	(0.096)	(0.046)
AGE3	−0.005	−0.159***	−0.189*	−0.186***
	(0.066)	(0.022)	(0.104)	(0.040)
TYPE1	−0.997***	−0.757***	−0.342	−0.247
	(0.348)	(0.128)	(0.700)	(0.252)
TYPE2	−1.467***	−1.105***	−0.628	−0.289
	(0.341)	(0.125)	(0.691)	(0.248)
SIZE1	0.706***	0.032	0.158	0.020
	(0.288)	(0.100)	(0.575)	(0.190)
SIZE2	0.497*	0.037	−0.216	0.051
	(0.287)	(0.100)	(0.579)	(0.188)
SIZE3	0.676**	0.079	−0.558	0.029
	(0.308)	(0.108)	(0.603)	(0.201)
CLASS1	0.362	0.649***	1.397***	0.877***
	(0.266)	(0.093)	(0.572)	(0.224)
CLASS2	0.717***	1.057***	0.861**	0.588***
	(0.207)	(0.080)	(0.404)	(0.171)
CLASS3	0.782***	1.112***	0.735*	0.516***
	(0.196)	(0.076)	(0.389)	(0.164)
CLASS4	0.871***	1.098***	0.658*	0.535***
	(0.193)	(0.076)	(0.383)	(0.162)
CLASS5	0.794***	1.008***	0.608	0.431***
	(0.196)	(0.077)	(0.384)	(0.163)
CLASS6	0.505***	0.901***	0.734*	0.354**
	(0.210)	(0.083)	(0.404)	(0.170)
CLASS7	0.367*	0.958***	0.636	0.267
	(0.217)	(0.093)	(0.415)	(0.180)
CLASS8	0.234	0.969***	0.629	0.414**
	(0.231)	(0.096)	(0.440)	(0.186)
Constant	−0.192*	0.158***	0.580***	1.337***
	(0.100)	(0.040)	(0.165)	(0.077)
REGION1	−0.188	0.053	0.397*	0.218**
(Kagoshima)	(0.125)	(0.047)	(0.222)	(0.099)
REGION2	−0.042	−0.004	−0.218	−0.204***
(Kanagawa)	(0.097)	(0.037)	(0.158)	(0.073)
REGION3	0.028	−0.057	−0.183	−0.256***
(Osaka)	(0.096)	(0.037)	(0.156)	(0.073)
N	2,813	19,184	1,101	10,405
$LogL$	−1818.903	−12901.859	−713.619	−3424.086

Note: ***, **, * indicate significance at a 1%, 5%, and 10% level, respectively. Standard errors are in the parentheses.

References

[1] Automobile Insurance Rating Organization (AIRO) (1999), "Jidosya hoken no gaikyo," (Survey of automobile insurance) in Japanese.

[2] Automobile Insurance Rating Organization (AIRO) (2002), "Automobile Insurance in Japan."

[3] Ausubel, L.M. (1999), "Adverse Selection in the Credit Card Market," Working Paper, Department of Economics, University of Maryland, 17 June 1999 http://www.ausubel.com/creditcard-papers/adverse.pdf/

[4] Barrett, G.F. and R. Conlon (2003), "Adverse Selection and the Decline in Private Health Insurance Coverage in Australia: 1989-95," *The Economic Record,* Vol. 79, No. 246, 279-296.

[5] Bond, E.W. and K.J. Crocker (1991), "Smoking, Skydiving, and Knitting: The Endogenous Categorization of Risks in Insurance Markets with Asymmetric Information, !I*Journal of Political Economy,* Vol. 99 (1), 177-200.

[6] Buchmueller, T. and J. DiNardo (2002), "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut," *American Economic Review,* Vol. 92 (1), 280-294.

[7] Cawley, J. and T. Philipson (1999), "An Empirical Examination of Information Barriers to Trade in Insurance," *American Economic Review,* Vol. 89, 827-46.

[8] Chiappori, P.A. and B. Salanié (2000a), "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy,* Vol. 108 (1), 56-78.

[9] Chiappori, P.A. and B. Salanié (2000b), "Testing Contract Theory: A Survey of some Recent Works," *Forthcoming, Advances in Economic Theory.*

[10] Chiappori, P.A., B. Julien, B. Salanié, and F. Salanié (2002), "Asymmetric Information in Insurance: Some Testable Implications," *Mimeo.* http://www.src.uchicago.edu/ pach/cjssdef.pdf/

[11] Cohen, A. (2002), "Asymmetric Information and Learning in the Automobile Insurance Market," *Harvard John M. Olin Discussion Paper Series No.371.*

[12] Crocker, K.J. and A. Snow (1986), "The Efficiency Effects of Categorical Discrimination in the Insurance Industry, !I*Journal of Political Economy*, Vol. 94 (2), 321-344.

[13] Crocker, K.J. and A. Snow (2000), "The Theory of Risk Classification," In *Handbook of Insurance* edited by Georges Dionne, Kluwer Academic Publishers.

[14] Dahlby, B.G. (1983), "Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance," *Journal of Public Economics,* Vol.20 (1), 121-130.

[15] Dahlby, B.G. (1992), "Testing for Asymmetric Information in Canadian Automobile Insurance," In *Contributions to Insurance Economics* edited by Georges Dionne, Kluwer Academic Publishers.

[16] Dionne, G. (2002), "Insurance Regulation in Other Industrial Countries," In *Deregulating Property-Liability Insurance,* ed. J. D. Cummins, 361-96 AEI-Brookings Join Center for Regulatory Studies.

[17] Dionne, G., C. Gouriéroux, and C. Vanasse (2001), "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment," *Journal of Political Economy,* Vol. 109, 444-453.

[18] Finkelstein, A. and J. Poterba (2000), "Adverse Selection in Insurance Markets: Policyholder Evidence from the UK Annuity Market," *NBER Working Paper,* W8045.

[19] Harrington, S.E. (1992), "Rate Suppression," *Journal of Risk and Insurance,* Vol. 56, pp.185-202.

[20] Harrington, S.E. and P.M. Danzon (2000), "Rate Regulation, Safety Incentives, and Loss Growth in Workers' Compensation Insurance," *Journal of Business,* Vol. 73 (4), 569-595.

[21] Hayakawa, H., P.S. Fischbeck, and B. Fischhoff (2002), "The Japanese Automobile Insurance Industry—Regulation, Market Structure and Possible Impacts of Deregulation, !I*Journal of Insurance Regulation,* Vol. 18, 385-403.

[22] Hemenway, D. (1990), "Propitious Selection," *Quarterly Journal of Economics,* Vol. 105 (4), 1063-69.

[23] Hemenway, David. (1992), "Propitious Selection in Insurance," *Journal of Risk and Uncertainty*, 5(3) pp.247-51.

[24] Hoy, Michael, (1982), "Categorizing Risks in the Insurance Industry," *Quarterly Journal of Economics,* pp.321-336.

[25] Institute for Traffic Accident Research and Data Analysis (ITARDA) (1999), "Kotsu-jiko Tokei Nenpo," (Annual Report on Traffic Accidents), in Japanese.

[26] Jaffee, D. M. and T. Russell (2002), "The Causes and Consequences of Rate Regulation in the Auto Insurance Industry," In *Deregulating Property-Liability Insurance,* ed. J. D. Cummins, 81-112, AEI-Brookings Join Center for Regulatory Studies.

[27] Jullien, B., B. Salanié, and F. Salanié (1999), "Should More Risk-Averse Agents Exert More Effort?," *The GENEVA PAPERS on Risk and Insurance Theory,* Vol. 24 (1), 19-28.

[28] Löfgren, K.G., T. Persson, and W.W. Jörgen (2002), "Markets with Asymmetric Information: The Contributions of George Akerlof, Michael Spence and Joseph Stiglitz," *Scandinavian Journal of Economics,* Vol. 104 (2), 195-211.

[29] National Police Agency, (1999), “Criminal Statistics,” annual.

[30] 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, 289-301.

[31] Rea, S. A. Jr. (1992), “Insurance Classifications and Social Welfare,” In *Contributions to Insurance Economics,* ed. George Dionne, 377-96. Boston: Kluwer.

[32] Saito, K. (2003), “Adverse Selection in the Early Automobile Insurance Market,” *Mimeo.*

[33] Siegelman, P. (2003), “Adverse Selection in Insurance Markets: An Exaggerated Threat,” *Fordham University School of Law Research Paper 27.*

[34] Statistical Bereau in Japan (1999), “Nihon no Tokei,” annual.