© *The Journal of Risk and Insurance*, 2006, Vol. 73, No. 2, 335-356

Testing for Asymmetric Information in the Automobile Insurance Market Under Rate Regulation

Kuniyoshi Saito

Abstract

This article examines whether adverse selection or moral hazard could be induced by rate regulation, which prohibits insurance companies from considering some attributes of drivers in setting premiums. Using an individual data set from a heavily regulated automobile insurance market, we arrived at several conclusions, as follows. First, no evidence of adverse selection or moral hazard is found in general: conditional on all the variables observed by insurer, the null hypothesis of independence between risk and coverage is not rejected at reasonable levels of statistical significance. Second, this result is robust in the sense that it holds under several empirical procedures and different definitions of risk and coverage. Third, we find that unobserved variables do not induce adverse selection: the null hypothesis that consumers in risky regions are more likely to purchase insurance is tested against the alternative and rejected. Our study supports the view that the adverse selection phenomenon exists only to a very limited extent in this market.

Introduction

Recent empirical studies in the competitive automobile insurance market show no signs of adverse selection in these markets. To take some of the outstanding studies, Chiappori and Salanié (2000) find no systematic relationship between risk and coverage in the French automobile insurance market, and Dionne, Gouriéroux, and

---

Kuniyoshi Saito works at the Department of Social Engineering, Tokyo Institute of Technology, O-okayama, Meguro-ku, Tokyo 152-8552, Japan. The author can be contacted via e-mail: ksaito@soc.titech.ac.jp. For helpful comments, I would like to thank Yoshiro Miwa, Hidehiko Ichimura, Michihiro Kandori, Yoshitsugu Kanemoto, Nobuhiro Kiyotaki, Motonari Kurasawa, Toshihiro Matsumura, Johannes Moenius, Dan Sasaki, Daisuke Tsuruta, Noriyuki Yanagawa, two anonymous referees, and the seminar participants at the University of Tokyo, as well as the conference participants at the 2003 Fall Meetings of the Japanese Economic Association. I am also grateful to a researcher at an insurance company for providing the data set and helpful comments and discussions. The earlier version of this article corresponds to the old manuscript titled "Does Less Risk Classification Induce more Adverse Selection?: Testing for Asymmetric Information in Auto Insurance Market." Any errors are, of course, my own responsibility.

Vanasse (2001) reach a similar conclusion using a data set in Quebec.[1] These studies use data from the competitive insurance market and their empirical results suggest that insurers are successfully managing adverse selection or moral hazard, at least in the competitive insurance market where they can freely set premiums.

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

These types of rate regulation have indeed succeeded in providing more coverage for high-risk agents who would otherwise be offered insurance with premiums too high to be purchased. However, there is a possibility that these regulations can induce serious adverse selection or moral hazard because premium discrimination has been considered to be one of the methods of mitigating these problems.[3]

In spite of the wide prevalence of rate regulation in the automobile insurance market, there have been few empirical studies to evaluate the impact of rate regulation on adverse selection or moral hazard.[4]

The purpose of this article is to examine whether adverse selection or moral hazard could be induced by these types of rate regulations. To this end, we use a data set on individuals from the Japanese automobile insurance market, where insurance rates have long been strictly regulated.

Our study differs from and improves on earlier works in two more ways. First, we test asymmetric information in various aspects. Since automobile insurance includes various types of coverage, it is possible that asymmetric information matters in some policies but not in others. For example, adverse selection could matter in the case of

---

[1] One exception is Cohen (2005), who finds a strong relationship between risk and coverage for drivers who have more than three years' driving experience before contracting with the insurer. Although her study suggests the importance of "learning" in this market, it should be noted that she finds such a relationship only for policyholders with three or more years of driving experience prior to joining the insurer. Generally speaking, more and more studies propose negative evidence of adverse selection or moral hazard in the automobile insurance market. See Chiappori and Salanié (2003), Löfgren, Persson, and Jörgen (2002), and Siegelman (2004), for broader surveys.

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

[3] For instance, Bond and Crocker (1991) state that "modest problems of adverse selection can be self-correcting if the market participants are permitted to use the categorization tools at their disposal." (p. 198) See Bond and Crocker (1991), Crocker and Snow (1986), Crocker and Snow (2000), Dahlby (1983), Hoy (1982), Jaffee and Russell (2002), Puelz and Kemmsies (1993), and Rea (1992), for the arguments regarding risk classification in insurance markets.

[4] One exception is Dahlby (1983), who examined the effects of prohibition of rate discrimination according to gender and showed that it induced adverse selection. Our study differs from his study in at least two aspects. One is that we use individual data, whereas Dahlby (1983) used aggregated data. The other is that we use the actual claim data, whereas Dahlby's results are based on the simulation method.

insurance for one's own vehicle damage, but not in the case of insurance for bodily injury liability. Considering this aspect, we use several definitions of "coverage" and "risk" to check the robustness of our basic results. Secondly, we examine the effects of rate prohibition *directly*. Suppose there is no risk classification by region and the traditional adverse selection story holds. Then consumers living in high-risk regions should be more likely to purchase insurance contracts than those in regions with fewer accidents. We examine whether this hypothesis is supported by our data.

Main findings can be summarized as follows. First, we find no evidence of adverse selection or moral hazard in general: conditional on all the variables observed by insurer, the null hypothesis of independence between risk and coverage is not rejected for most of the groups. Rather, we find *negative* correlation between risk and coverage for some groups, which is the reverse of the adverse selection hypothesis. Second, the result is robust in the sense that it holds under several empirical procedures and definitions of coverage and risk. Third, our study shows that variables that are observed by the insurer but not used for premium discrimination do not necessarily induce adverse selection: the hypothesis that the consumers living in high accident probability regions are likely to purchase more insurance is tested against the alternative and is rejected. From these findings, we believe that adverse selection or moral hazard phenomena exist only to a limited extent in this market and can be mitigated by some basic mechanisms such as deductibles and experience ratings.

The remainder of this article is organized as follows. In the next section, we provide a brief introduction of the Japanese automobile insurance market. In the subsequent section, we propose the empirical procedure. The subsection "Description of the Data" introduces our data set, and empirical results are reported in the next section. Further investigations are provided in the subsequent section. Concluding remarks are presented in the last section.

THE JAPANESE AUTOMOBILE INSURANCE MARKET

In Japan, automobile insurance is operated under two different systems, compulsory automobile liability insurance and voluntary insurance.[5] Compulsory insurance only covers liability for bodily injury, and the law specifies the limit of liability for death and each grade of permanent disability and bodily injury. On the other hand, voluntary insurance includes third-party liability coverage (bodily injury liability and property damage liability), self-incurred personal accident coverage, and own vehicle damage coverage. With respect to liability for bodily injury, voluntary insurance acts as excess cover for compulsory insurance.

Prior to 1998, premium rates had been strictly regulated in this market. There was a nonprofit organization called the "Automobile Insurance Rating Organization" (AIRO). AIRO calculated both compulsory and voluntary insurance premium rates using the data gathered from its member insurers. The premium rates were filed with the Commissioner of the Financial Services Agency (FSA), and were examined on the basis of three principles of premium rates, which were: "reasonable," "adequate,"

[5] The description of the Japanese automobile insurance market in this section is based on Automobile Insurance Rating Organization (2002). See Dionne (2002), and Hayakawa, Fischbeck, and Fischhoff (2002) for more information about this market.

TABLE 1
Risk Segmentation in Japan (Prior to and After 1998)

Variable	Prior to 1998	After 1998
Age	Three categories: All ages, 21 years or over, 26 years or over	Any number of categories allowed, but differentials between the highest and lowest rated groups to be within a range of 300% or less
Gender	No segmentation	Segmentation allowed, but differentials between male and female to be within a range of 150% or less
Driving history	Segmentation depends on number of no-claims years. Accident-free drivers can qualify for a discount of up to 60% under the base rate. Drivers with a history of accidents have to pay up to 50% over the base rate	Any kind of segmentation allowed. (However, as driving records are currently not publicly available, insurers cannot introduce policyholders' driving records as a risk factor.)
Auto usage	No segmentation	Usage such as commercial, personal, leisure, commuting, and so on
Pattern of use	No segmentation	For example, mileage per year
Geography	No segmentation	Maximum division up to seven regions. Differentials between regions to be within a range of 150% or less
Vehicle type	Passenger automobile, Truck, Bus, Motorcycle, Dump truck, Other (specific automobile)	Any kind of segmentation allowed
Multicar ownership	No segmentation	Discount may be applied based on the number of vehicles insured
Vehicle safety features	Discount is applied based on presence of safety features such as airbags	Discount may be applied depending on presence of safety features such as airbags, anti-lock brake systems, and so on

Source: Automobile Insurance Rating Organization (2002).

and "not unfairly discriminatory." Most insurance companies joined AIRO, and it was forced to make its members abide by the AIRO rates by law.[6] As a result, almost all the consumers had no choice but to purchase contracts calculated by AIRO.

Table 1 provides detailed information about the risk segmentation allowed prior to and after the deregulation in 1998. It shows that, prior to partial deregulation, there had

[6] Although joining the AIRO was not compulsory, most insurers became members.

been no discrimination by gender, auto usage, pattern of use, geography, or multicar ownership, all of which seem to have strong correlations with the driver's probability of having an accident.

There is little doubt that rate regulation had been a binding constraint for insurance companies because of the fact that more and more insurance companies offered new contracts that reflected this deregulation after the partial rate liberalization in 1998. Therefore, the Japanese automobile insurance market provides an ideal situation for examining the effects of rate regulation on adverse selection or moral hazard.[7]

EMPIRICAL PROCEDURE

One of the central shared predictions in the economics of information is that, in the presence of asymmetric information, high-risk agents are more likely to purchase insurance: under the condition that insurers cannot offer the contracts reflecting each agent's risk level, high-risk agents are more likely to purchase insurance because the marginal utility is higher for them.

Chiappori et al. (2002) show that this property holds in general, and they argue that the test of adverse selection or moral hazard can be simply expressed as a conditional correlation test between coverage and risk.[8] The test should be "conditional" because the relationship between y_i and z_i only holds within a group of drivers that insurance companies regard as the "same."[9] We basically follow the empirical approach of Chiappori and Salanié (2000), using several definitions of coverage and risk.

Let us denote the coverage by y_i. Since automobile insurance includes various types of coverage, we define y_i in the following two ways.

$$y_i^{\text{Vehicle}} = \begin{cases} 1 & \text{if agent } i \text{ purchased coverage for own vehicle damage,} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_i^{\text{Ded}} = \begin{cases} 1 & \text{if agent } i \text{ purchased coverage for own vehicle with zero deductibles,} \\ 0 & \text{otherwise.} \end{cases}$$

[7] Although our data cover the period from 1999 to 2000, it was several years after the deregulation that these insurance companies began to offer new contracts reflecting partial rate liberalization. Therefore, our data are still valid for the purpose of this study.

[8] Chiappori et al. (2002) argue that this property basically holds in a competitive market (or more precisely, the zero-profit condition). Since the automobile insurance market in Japan is strictly regulated as we described in the section "The Japanese Automobile Insurance Market," we have to check whether it is appropriate to apply this property to the Japanese automobile insurance market. As Chiappori et al. (2002) argue, "the zero-profit condition" can be weakened to the condition that the contracts with broader coverage have (weakly) less profit. Then we checked whether this condition holds in this market and found the premium:loss ratio is consistently higher for contracts with broader coverage, which suggests that contracts with broader coverage have less profit.

[9] A conditional correlation test does not tell anything about whether the correlation comes from adverse selection or moral hazard. Although disentangling these is a recent empirical issue, we do not ask whether it comes from adverse selection or moral hazard.

We focus on coverage for own vehicle damage because whether or not to purchase this contract is an important decision for consumers in Japan. The premium is much higher for a contract with this coverage: In 1997, the average annual premium of a typical contract with and without coverage for own vehicle damage is 85,750 yen (US $715) and 51,247 yen (US $427), respectively. The premium also largely varies according to the level of deductibles. Therefore, the choice between these types of coverage is an important decision.

There are two cases for $y_i^{\text{Ded}} = 1$. One is that the policyholder purchased a 0–10 proportional contract, which means that the deductibles are zero for the first claim and 100,000 yen for the second claim in the same policy year. The other case is that the policyholder purchases a 5–10 proportional contract and also chooses the zero-deductible option, which sets the deductible to zero only for the first claim.

We define risk z_i in the following two ways.

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

$$z_i^{\text{Theft}} = \begin{cases} 1 & \text{if agent } i\text{'s car was stolen in the policy year} \\ 0 & \text{otherwise.} \end{cases}$$

In the definitions of risk, the "accident $\neq$ claim" problem described below is quite important. Consider the case where agent i's car was stolen. If he/she had purchased an insurance contract for his/her own vehicle damage, he/she might claim from the insurance company. However, if he/she had not purchased that contract, no claim would be filed even though his/her car was stolen. Since our data come from an insurance company and we can only observe a claim, not an accident, we are likely to suffer from this bias if we do not appropriately control for it.

To avoid this "accident $\neq$ claim" problem, we define risk in such a way that an accident will almost always become a claim regardless of the contract the agent purchased. In the first equation, risk is defined as whether or not the agent i had at least one crash in the policy year. Since the damage to the other vehicle in a crash is covered by the property damage liability insurance contract, which is purchased by most policyholders with voluntary insurance, the claim will be filed in the insurer's file whether or not he/she purchased coverage for his/her own vehicle damage.

Using these definitions of coverage and risk, we can make $2 \times 2 = 4$ tests on the conditional correlation between y_i and z_i. Among these four combinations, however, one test ($(y_i, z_i) = (y_i^{\text{Vehicle}}, z_i^{\text{Theft}})$) cannot be tested because of the "accident $\neq$ claim" problem described above. Therefore, we conduct three tests on the conditional independence between y_i and z_i.

DESCRIPTION OF THE DATA

The Data

Our data are from a large insurance company in Japan. The original data set includes 30,000 voluntary insurance policies undertaken during the period from April 1, 1999 to March 31, 2000.

Since the partial liberalization took place in 1998, it was already permitted for insurers to use additional variables at the time of data extraction. However, as we noted in the section "The Japanese Automobile Insurance Market," it was several years after the deregulation that most of the large insurance companies began to offer new contracts reflecting the deregulation. The insurance company from which we collected our data set was also selling the same contract as before the deregulation at the time of data extraction, and there was no discrimination by gender, auto usage, pattern of use, geography, or multicar ownership.[10] Therefore, this data set is ideal for our purposes.

The data were extracted using stratified sampling. In 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) location in the east or west of Japan. From these categories, four representative prefectures were selected: Iwate, Kanagawa, Osaka, and Kagoshima.[11] The characteristics of each prefecture are summarized in Table 2.

In the second step, using the prefecture-level diffusion rate and average claim rate of collision insurance, we calculated the number of contracts that were expected to include a sufficient number of accidents (30,000). With this total number of contracts fixed, the data were extracted from each prefecture proportional to the number of voluntary insurance contracts. These two steps yielded the number of contracts in each area as follows: Kanagawa (12,246), Osaka (12,829), Iwate (2,152), Kagoshima (2,773), Total (30,000).[12]

Since the data set includes several vehicle types (truck, bus, motorcycle, etc.), we limited our sample to private cars. There are 21,997 such contracts. In addition, we divided the data into "beginners" and "experienced drivers" for the following two reasons. First, the accuracy of risk perception could be different according to driving experience. A driver might have more accurate knowledge about his own accident probability when he/she has more driving experience. Secondly, a policyholder may not be able to evaluate the contents of the contract for his first purchase, whereas a driver with more experience could be more knowledgeable about the contracts he/she purchases. In either case, drivers' actions could be different according to their driving experience.

[10] Several years later, this company also began to offer contracts that reflected the partial rate liberalization.

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

[12] Three caveats are worth mentioning. First, 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. Second, coinsurance contracts are excluded. Policies are sometimes undertaken by several insurers, an arrangement called coinsurance. Since such contracts made our study complicated, we excluded such contracts and limited the study to the policies undertaken by one insurer. Finally, almost all the accidents are captured in our data and there are few cases when accidents are incurred but not reported. It sometimes takes several years for an accident to be settled in the data file, especially in cases of accidents involving bodily injury. Since the policies were contracted during the policy year of 1999 and the data were extracted in February 2003, almost all the claims are fixed at the time of data extraction.

TABLE 2
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

Notes: All the figures are in 1999 values. Population density is measured in persons per square kilometers.
Source: Automobile Insurance Rating Organization (1999) ("Diffusion rate" and "Claim rate") and Statistical Bureau of Japan (1999) ("Population Density").

Although it would be ideal to divide the data set into beginners and experienced drivers using information such as the age of driver or the year of obtaining the driver's license, our data do not include those variables. Therefore, we used experience rating class as a proxy for the driver's experience. A policyholder belongs to class 6 for the first year, going down three classes if he/she makes a claim, and one class up if he/she makes no claim.[13] This system suggests that most of the policyholders in class 6 or 7 are those who have less than two years driving experience.[14]

Using the experience rating class as a proxy for the driver's experience, we call the policyholders who belong to classes 6 or 7 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.

Description of the Data

To begin with, let us give an overview of our data set. The findings in this subsection should be viewed only as descriptive statistics.

Table 3 shows the cross-tabulations under several definitions of y_i and z_i.[15] The three tables in the first column show the cross-tabulations using the definition of (y_i^{Vehicle}, z_i^{Crash}). Out of 21,997 policies, 52.31 percent purchased insurance for own vehicle damage. The percentage of drivers who make claims is slightly higher for those with that type of insurance: 5.34 percent for $y_i^{\text{Vehicle}} = 1$ and 4.95 percent for $y_i^{\text{Vehicle}} = 0$, which is consistent with the existence of adverse selection or moral hazard.

[13] This insurance company uses 1 to 18 experience rating classes.
[14] Needless to say, there are some drivers who once belonged to other classes and then came back to class 6 or 7 at the time of data extraction. However, according to the insurer, the number of those policyholders is not large.
[15] Second and third claims are ignored throughout this article. However, the percentage of policies which have second or third claims is low. For example, out of 1,133 claims in the cross-tabulation using the definition of (y_i^{Vehicle}, z_i^{Crash}) in Table 3, only twenty-six claims (2.29 percent) have second or third claims.

TABLE 3
Cross-Tabulations

A. $(y_i, z_i) = (y_i^{\text{Vehicle}}, z_i^{\text{Crash}})$

Beginners

y_i^{Vehicle}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.641$

Experienced Drivers

y_i^{Vehicle}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.094$

Total

y_i^{Vehicle}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.192$

B. $(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Crash}})$

Beginners

y_i^{Ded}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.276$

Experienced Drivers

y_i^{Ded}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.040$

Total

y_i^{Ded}	z_i^{Crash} 0	z_i^{Crash} 1	Total	Claim (%)
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 $= 0.003$

C. $(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Theft}})$

Beginners

y_i^{Ded}	z_i^{Theft} 0	z_i^{Theft} 1	Total	Claim (%)
0	446	4	450	0.89
	40.84	44.44	40.87	
1	646	5	651	0.77
	59.16	55.56	59.13	
Total	1,092	9	1,101	0.82
	100	100	100	

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

Experienced Drivers

y_i^{Ded}	z_i^{Theft} 0	z_i^{Theft} 1	Total	Claim (%)
0	1,093	7	1,100	0.64
	10.56	13.46	10.57	
1	9,260	45	9,305	0.48
	89.44	86.54	89.43	
Total	10,353	52	10,405	0.50
	100	100	100	

$\chi^2 = 0.462$ Pr $= 0.497$

Total

y_i^{Ded}	z_i^{Theft} 0	z_i^{Theft} 1	Total	Claim (%)
0	1,539	11	1,550	0.71
	13.45	18.03	13.47	
1	9,906	50	9,956	0.50
	86.55	81.97	86.53	
Total	11,445	61	11,506	0.53
	100	100	100	

$\chi^2 = 1.095$ Pr $= 0.295$

The same trend can be found for both beginners and experienced drivers. For beginners (experienced drivers), the percentage of claims is 6.99 percent (5.16 percent) for $y_i^{\text{Vehicle}} = 1$ drivers, compared with 6.54 percent (4.63 percent) for y_i^{Vehicle} drivers. This association between y_i and z_i is not statistically significant (Pr = 0.641) for beginners, but is statistically significant (Pr = 0.094) for experienced drivers.

Three tables in the middle column of Table 3 show that the tendency is slightly different when we define coverage as whether or not a driver purchases a policy with a positive amount of deductibles (y_i^{Ded}). Among 21,997 policyholders, 11,506 drivers purchase insurance for own vehicle damage. Out of 11,506 policies, 9,956 policies (86.53 percent) are with collision insurance with zero-deductible. For beginners (experienced drivers), the percentage of claim is 6.30 percent (5.01 percent) for $y_i^{\text{Ded}} = 1$ group, and 8.00 percent (6.45 percent) for $y_i^{\text{Ded}} = 0$ group. This association is comparatively strong, especially for experienced drivers (Pr = 0.040).[16]

In the third column of Table 3, we show the cross-tabulations using the definitions of y_i^{Ded} and z_i^{Theft}. They show that the percentage of claim is slightly lower for the policyholders with zero-deductible coverage, but the χ^2 statistic shows that this association is not statistically significant for beginners or experienced drivers.

To summarize, a first look at our data set shows that: (1) we find a positive but weak correlation between y_i^{Vehicle} and z_i^{Crash}; (2) we find a negative correlation between y_i^{Ded} and z_i^{Crash}, which is statistically significant at 10 percent only for experienced drivers; and (3) we find a negative correlation between y_i^{Ded} and z_i^{Theft}, but the relationship is not statistically significant at all.

ESTIMATION RESULTS

A simple investigation of our data so far suggests that, consistent with the story of adverse selection or moral hazard, the relationship between y_i and z_i is positive when we define coverage and risk as y_i^{Vehicle} and z_i^{Crash}, respectively. In order to test asymmetric information, however, we have to control for all the variables observed by the insurer, because the independence of risk and coverage should be tested *within* the group of drivers that the insurer regards as belonging to the same risk class.

We denote as X_i the set of variables that the insurance company observes about each policyholder. Four variables are included in X_i: range of age covered by the policy (4), type of vehicle (3), size of vehicle (5), and class of insurance for own vehicle damage (10)[$(\cdot)$ denotes the number of classifications].[17] Definitions of these variables and the number of policyholders in each category are described in Table 4.

To test the conditional independence between y_i and z_i, we conduct two empirical procedures following Chiappori and Salanié (2000): bivariate probit modeling and a χ^2 test.

[16] This result is consistent with 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. We leave the discussion of this story to future research.

[17] Family restriction discount and safety device discount also affect the premium. These variables are not included because the data were not available.

TABLE 4
Number of Beginners and Experienced Drivers According to Classes

		Beginners		Experienced Drivers	
Variables	Definition	Count	Percentage	Count	Percentage
AGE1	=1 if the policy covers all ages	330	11.73	1,137	5.93
AGE2	=1 if the policy covers those 21 years of age or over	831	29.54	3,197	16.66
AGE3	=1 if the policy covers those 26 years of age or over	625	22.22	4,902	25.55
AGE4	=1 if the policy covers those 30 years of age or over	1,027	36.51	9,948	51.86
TYPE1	=1 if the policyholder's vehicle is standard-size passenger automobile	649	23.07	5,289	27.57
TYPE2	=1 if the policyholder's vehicle is small-size passenger automobile	1,411	50.16	10,559	55.04
TYPE3	=1 if the policyholder's vehicle is light-size passenger automobile	753	26.77	3,336	17.39
SIZE1	=1 if the horse power is under 1500 cc	530	18.84	3,973	20.71
SIZE2	=1 if the horse power is under 2500 cc	1,196	42.52	9,623	50.16
SIZE3	=1 if the horse power is over 2501 cc	306	10.88	2,088	10.88
SIZE4	=1 if the vehicle is of special use	28	1.00	164	0.85
SIZE5	=1 if no classification by horse power	753	26.77	3,336	17.39
CLASS1	=1 if the class of insurance for own vehicle damage is 1	54	1.92	527	2.75
CLASS2	=1 if the class of insurance for own vehicle damage is 2	217	7.71	1,797	9.37
CLASS3	=1 if the class of insurance for own vehicle damage is 3	564	20.05	5,004	26.08
CLASS4	=1 if the class of insurance for own vehicle damage is 4	473	16.81	3,857	20.11
CLASS5	=1 if the class of insurance for own vehicle damage is 5	346	12.30	2,518	13.13
CLASS6	=1 if the class of insurance for own vehicle damage is 6	139	4.94	886	4.62
CLASS7	=1 if the class of insurance for own vehicle damage is 7	116	4.12	457	2.38
CLASS8	=1 if the class of insurance for own vehicle damage is 8	85	3.02	405	2.11
CLASS9	=1 if the class of insurance for own vehicle damage is 9	34	1.21	173	0.90
CLASS10	=1 if no classification by insurance for own vehicle damage	785	27.91	3,560	18.56

Bivariate Probit

We define two probit models as follows,

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

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

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: $\mathrm{E}[\epsilon_i] = \mathrm{E}[\eta_i] = 0$, $\mathrm{Var}[\epsilon_i] = \mathrm{Var}[\eta_i] = 1$, and $\mathrm{Cov}[\epsilon_i, \eta_i] = \rho$. In this model, ρ measures the correlation between coverage and risk after the influence of the included observations is accounted for. We estimate this model using several definitions of y_i and z_i as before.

Table 5 presents the estimation results of the bivariate probit modeling. These tables suggest that the findings in the previous section are not altered significantly even if we control for the variables observed by insurer.

The first two estimation results show that we again find positive correlation between coverage and risk when they are defined as y_i^{Vehicle} and z_i^{Crash}, respectively. The estimated values of ρ are 0.062 for beginners and 0.031 for experienced drivers. The p-values of ρ, however, suggest that this association is not statistically significant. Therefore, we find no evidence of adverse selection or moral hazard when the observed variables are taken into account.

The next two estimation results suggest that the correlation between coverage and risk is *negative* when they are defined as y_i^{Ded} and z_i^{Crash}. The estimated values of ρ are -0.076 for beginners and -0.065 for experienced drivers. This association is not strong and is significant only for experienced drivers with a p-value of 0.034.[18] Therefore, our estimation results show no evidence of asymmetric information when we focus on the level of deductibles.

Finally, the estimation results on the right-hand side of Table 5 show that we again find *negative* correlation between y_i and z_i, when they are defined as y_i^{Ded} and z_i^{Theft}.[19]

[18] Although we do not yet have a compelling explanation for this finding, one possible explanation would be heterogeneity in risk perceptions. If the agents who inappropriately assess their risk to be low tend to choose a deductible or coverage consistent with their low assessment of risk, and if they tend to be less willing to take precautions, the relationship between coverage and *ex post* risk could exhibit negative or no correlation (see, e.g., Koufopoulos, 2002). Recent study also indicates that biological, psychological, and behavioral characteristics could provide an explication of the relationship between financial decision making and risky driving habits (Brockett and Golden, 2005). However, the finding is still puzzling since negative correlation is found for "experienced" drivers who have learned and are well informed.

[19] We do not divide the data into beginners and experienced drivers because the number of thefts is too small (nine) in the case of beginners.

TABLE 5
Bivariate Probit Estimation Results

Independent Variables	$(y_i, z_i) = (y_i^{\text{Vehicle}}, z_i^{\text{Crash}})$				$(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Crash}})$				$(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Theft}})$	
	Beginners		Experienced Drivers		Beginners		Experienced Drivers		Total	
	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Choice of contract equation										
AGE1	−0.385***	0.000	−0.306***	0.000	−0.191	0.221	−0.178**	0.019	−0.253***	0.000
AGE2	0.032	0.591	−0.204***	0.000	−0.434***	0.000	−0.318***	0.000	−0.447***	0.000
AGE3	0.001	0.983	−0.161***	0.000	−0.205**	0.047	−0.193***	0.000	−0.215***	0.000
TYPE1	−6.762	0.999	−6.528	0.997	1.005	0.166	−0.293	0.245	−0.296	0.196
TYPE2	−7.237	0.999	−6.873	0.997	0.742	0.299	−0.326	0.190	−0.369	0.103
SIZE1	0.678**	0.019	0.040	0.691	0.171	0.765	0.037	0.847	0.016	0.927
SIZE2	0.456	0.114	0.032	0.748	−0.180	0.754	0.057	0.764	0.014	0.935
SIZE3	0.610**	0.049	0.026	0.812	−0.498	0.406	0.026	0.896	−0.073	0.695
CLASS1	6.153	0.999	6.420	0.997	—	—	0.899***	0.000	1.034***	0.000
CLASS2	6.514	0.999	6.827	0.997	−0.552	0.220	0.599***	0.000	0.698***	0.000
CLASS3	6.580	0.999	6.881	0.997	−0.697	0.109	0.514***	0.002	0.620***	0.000
CLASS4	6.674	0.999	6.873	0.997	−0.758*	0.080	0.533***	0.001	0.593***	0.000
CLASS5	6.601	0.999	6.787	0.997	−0.815*	0.063	0.436***	0.007	0.491***	0.001
CLASS6	6.320	0.999	6.697	0.997	−0.704	0.133	0.358**	0.035	0.440***	0.003
CLASS7	6.186	0.999	6.759	0.997	−0.822	0.088	0.270	0.134	0.318**	0.043
CLASS8	6.052	0.999	6.777	0.997	−0.853	0.090	0.413**	0.026	0.466***	0.004
CLASS9	6.273	0.999	6.516	0.997	−1.428**	0.011	1.179***	0.000	1.075***	0.000
Constant	−0.212***	0.000	0.139***	0.000	0.457***	0.000	—	—	—	—

(continued)

TABLE 5
(Continued)

	$(y_i, z_i) = (y_i^{\text{Vehicle}}, z_i^{\text{Crash}})$				$(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Crash}})$				$(y_i, z_i) = (y_i^{\text{Ded}}, z_i^{\text{Theft}})$	
	Beginners		Experienced Drivers		Beginners		Experienced Drivers		Total	
Independent Variables	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Accident (theft) equation										
AGE1	0.493***	0.000	0.207***	0.001	0.577***	0.003	0.169*	0.056	−0.275	0.402
AGE2	0.087	0.351	0.063	0.143	0.060	0.675	0.075	0.194	0.218*	0.077
AGE3	−0.127	0.246	−0.017	0.659	−0.152	0.369	−0.052	0.311	−0.002	0.984
TYPE1	0.190	0.689	−0.149	0.546	0.608	0.475	0.208	0.537	−2.675***	0.000
TYPE2	0.407	0.376	−0.183	0.452	0.803	0.333	0.171	0.608	−2.857	.
SIZE1	−0.571*	0.065	−0.136	0.392	−0.783	0.227	0.208	0.433	3.538***	0.000
SIZE2	−0.350	0.249	−0.102	0.517	−0.668	0.309	0.222	0.402	3.746***	0.000
SIZE3	−0.388	0.271	−0.185	0.278	−0.382	0.588	0.115	0.680	3.721***	0.000
CLASS1	−0.439	0.424	0.252	0.234	—	—	−0.543**	0.037	−4.791	0.999
CLASS2	0.151	0.703	0.344*	0.075	−0.103	0.854	−0.415**	0.049	−0.832**	0.020
CLASS3	0.182	0.632	0.369*	0.051	−0.035	0.947	−0.309	0.126	−1.028***	0.002
CLASS4	0.169	0.657	0.341*	0.072	−0.122	0.819	−0.366*	0.068	−0.656**	0.022
CLASS5	0.284	0.460	0.435**	0.023	0.131	0.808	−0.325	0.106	−0.486*	0.083
CLASS6	0.233	0.563	0.257	0.203	−0.027	0.964	−0.474**	0.028	−0.406	0.169
CLASS7	−0.016	0.969	0.397*	0.060	−0.580	0.406	−0.148	0.503	0.038	0.896
CLASS8	−0.188	0.682	0.417*	0.052	−0.110	0.866	−0.418*	0.079	0.058	0.843
CLASS9	−5.076	1.000	0.455*	0.061	−5.317	1.000	−1.676***	0.000	−2.941***	0.000
Constant	−1.652***	0.000	−1.731***	0.000	−1.551***	0.000	—	—	—	—
N	2,813		19,184		1,101		10,405		11,506	
ρ	0.062	0.198	0.031	0.119	−0.076	0.315	−0.065**	0.034	−0.046	0.521
LogL	−2479.448		−16577.992		−989.097		−5552.746		−4763.648	

Notes: In each of the equations, SIZE4 is dropped due to the problem of multicolinearity. In estimating the equation using the definitions of (y_i^{Ded}, z_i^{Theft}), we do not divide the data into beginners and experienced drivers because the number of thefts is too small (nine observations) in the case of beginners.

***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 6
Categorization Used for χ^2 Test

Variable	a	b
AGE	AGE1–3	AGE4
TYPE	TYPE1, TYPE3, TYPE4	TYPE2
SIZE	SIZE1, SIZE3–SIZE5	SIZE2
CLASS	CLASS4–CLASS10	CLASS1–CLASS3

Notes: Variables are defined as in Table 4. The criteria in this table are determined so as to have a sufficient number of samples in each cell.

However, this association is not statistically significant (p-value is 0.521). In summary, all these results show no evidence of adverse selection or moral hazard, even if we take into account the variables observed by insurer.

χ^2 Test

A 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 results so far, we test the conditional independence of y_i and z_i by χ^2 test.

First, we divide our data into several groups. Using the criteria shown in Table 6, the data are divided into $2^4 = 16$ cells. The criteria in Table 6 are defined so that there is a sufficient number of samples in each cell. For each cell, we construct cross-tabulations defined as in Table 3, and calculate χ^2 statistics, Q, which follows a $\chi^2(1)$ distribution.

We also calculate the statistics S, defined as the summation of Q, which also follows the χ^2 distribution under the null of independence between y_i and z_i.

Table 7 shows the estimated values of Q.[20] The left-hand side of the table shows the estimation results using the definition of (y_i^{Vehicle}, z_i^{Crash}). The values of Q are small in most of the cells and it is statistically significant only in one group: (b, a, a, a) of the experienced drivers.[21] When we further categorize this cell using additional variables in X_i, however, we find no statistically significant correlation. The S statistic also shows that the null hypothesis of independence between y_i^{Vehicle} and z_i^{Crash} is not rejected for the total sample.

[20] χ^2 test using the definitions of (y_i^{Ded}, z_i^{Theft}) is not executed because there are too few cases for $z_i^{\text{Theft}} = 1$.

[21] Some information about this group would be helpful. The number of policies with collision insurance is 986 out of the 2,116 in the total sample. The claim rate of $y_i^{\text{Vehicle}} = 1$ group is 6.39 percent (=63/986) and that of $y_i^{\text{Vehicle}} = 0$ group is 4.61 percent (=47/1,020). The correlation comes from the group of small-sized passenger automobiles, covered for drivers 21 years of age or over.

TABLE 7
χ^2 Test Results

	$(y_i^{\text{Vehicle}}, z_i^{\text{Crash}})$				$(y_i^{\text{Ded}}, z_i^{\text{Crash}})$			
	Beginners		Experienced Drivers		Beginners		Experienced Drivers	
Group	Q	N	Q	N	Q	N	Q	N
(a, a, a, a)	0.962	202	0.005	2,567	0.939	74	0.554	1,478
(b, a, a, a)	0.033	335	2.751*	2,116	0.160	113	0.207	1,049
(a, b, a, a)	0.82	64	0.159	1,016	1.126	34	1.862	682
(a, a, b, a)	1.714	75	0.226	839	0.157	25	0.021	410
(a, a, a, b)	0.178	203	1.836	2,047	1.548	80	3.262*	1,137
(b, b, a, a)	0.192	125	0.006	899	3.758*	62	0.662	548
(b, a, b, a)	0.103	241	0.601	903	0.013	59	0.691	334
(b, a, a, b)	0.268	329	0.065	1,992	0.054	130	2.349	961
(a, b, b, a)	0.016	66	0.392	709	1.193	35	0.014	493
(a, b, a, b)	—	20	0.007	291	—	10	0.495	173
(a, a, b, b)	—	8	0.635	54	—	0	—	9
(b, b, b, a)	0.455	88	0.059	574	2.002	41	1.305	311
(b, b, a, b)	0.599	30	0.313	257	0.950	19	2.079	147
(b, a, b, b)	0.865	18	0.36	41	—	5	—	6
(a, b, b, b)	1.975	389	2.626	2,425	0.313	165	0.577	1,394
(b, b, b, b)	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

Notes: Since there are no $y_i = 1$ or $z_i = 1$ cases for some cells, we cannot calculate Q statistics. Such cases are denoted by "—". An χ^2 test using the definitions of $(y_i^{\text{Ded}}, z_i^{\text{Theft}})$ was not conducted because the number of observations for $z_i^{\text{Theft}} = 1$ is small.
* indicates significance at the 10% level.

The right-hand side of Table 7 shows the estimation results using the definitions of $(y_i^{\text{Ded}}, z_i^{\text{Crash}})$. We find three Q values that are statistically significant: (b, b, a, a) for beginners, (a, a, b, b) and (b, b, b, b) for experienced drivers. These correlations do not disappear even if we include additional variables in X_i. It should, however, be noted that these correlations between y_i^{Ded} and z_i^{Crash} are *negative*, which suggests that the drivers with zero-deductible contracts are less likely to have accidents. The S statistics, however, show that those associations are not statistically significant as a whole.

To summarize, the χ^2 test procedures show that: (1) we find positive but not statistically significant correlation between y_i^{Vehicle} and z_i^{Crash} for some groups; (2) we find *negative* correlation between y_i^{Ded} and z_i^{Crash}, which is statistically significant at 10 percent for experienced drivers; and (3) we do not find any statistically significant correlation between y_i and z_i for the total sample.

These results show that the χ^2 tests also support our basic finding of no evidence of adverse selection or moral hazard.

DISCUSSIONS

We have defined risk as whether or not the policyholder has had at least one crash, but the severity of an accident has not been taken into account so far. To check the

TABLE 8
Average Amount of Loss

	No. of Claims	Mean	S.E.	Two-Sample Test	Rank-Sum Test
Beginners					
$y_i^{\text{Vehicle}} = 0$	112	927,618	177,058	1.616	0.395
$y_i^{\text{Vehicle}} = 1$	77	627,710	55,516	0.108	0.693
$y_i^{\text{Ded}} = 0$	36	456,155	78,434	0.031	−0.358
$y_i^{\text{Ded}} = 1$	41	453,058	61,962	0.975	0.721
Experienced Drivers					
$y_i^{\text{Vehicle}} = 0$	407	703,784	41,525	−0.399	1.117
$y_i^{\text{Vehicle}} = 1$	537	732,172	57,803	0.690	0.264
$y_i^{\text{Ded}} = 0$	71	817,388	189,478	2.556 **	0.437
$y_i^{\text{Ded}} = 1$	465	528,181	33,535	0.011	0.662

Notes: The null hypothesis for a two-sample test is that the means for $y_i^d = 1$ and $y_i^d = 0$ groups are equal. The null hypothesis for Wilcoxon's rank-sum test is that the medians for these groups are equal.
** indicates significance at the 5% level.

robustness of our results, we redefine risk as the amount of loss paid for crashes and examine whether there is a significant difference between the agents with different types of coverage.

Amount of Loss Paid

We define risk as follows:

$$z_i^{\text{Loss}} = \{\text{amount of loss paid for crashes}\}.$$

Table 8 reports the average amount of loss paid. In making this table, all the accidents with less than 50,000 yen for policyholders with zero-deductible contracts were discarded (five cases), and we added 50,000 yen for loss paid to those without zero-deductible contracts. These adjustments are due to the "accident $\neq$ claim" problem, which is described earlier. In addition, we discarded the contracts with losses of more than 10 million yen as outliers (three cases), and the second or third claims were ignored.

We can see that the average loss amount is higher for the $y_i = 1$ group in three out of four cases. In one case (experienced drivers ($y_i^{\text{Vehicle}}, z_i^{\text{Loss}}$)), the average loss paid is higher for policyholders with broader coverage. In this case, however, the difference is not large.

In order to check whether these differences are statistically significant, we conduct two empirical procedures: a two-sample test and Wilcoxon's rank-sum test.

The test results are also shown in Table 8. The two-sample test suggests that the mean difference is only statistically significant at the 5 percent level for one test: deductibles of experienced drivers. Note, again, that the relationship is *negative*, which contradicts

TABLE 9
Number of Crashes and Thefts in Each Prefecture

Prefecture	Crashes (A)	Thefts (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: The number of thefts is from the National Police Agency (1999). The numbers of crashes and private cars are from ITARDA (1999).

the adverse selection or moral hazard hypothesis. Wilcoxon's rank-sum test suggests, however, that this correlation is not significant by nonparametric procedure: The null hypothesis of equal medians for $y_i = 1$ and $y_i = 0$ group is not rejected in every test.

We also regressed the choice of contract y_i^{Vehicle} (or y_i^{Ded}) on all the attributes of the policyholders X_i and the amount of loss in logarithms. Estimation results by Probit modeling show that the estimated coefficients of the loss amount are not statistically significant in either of the models.[22]

To summarize, both the parametric and nonparametric empirical procedures show that positive correlation between the choice of deductibles and the amount of loss is not found either for beginners or experienced drivers.

Do Unobserved Variables Induce Adverse Selection?

As we have described in Table 1, premiums were not varied according to region until the rate liberalization in 1998. When we look at the data from the police department, however, we can see that the probabilities of accidents and theft vary considerably among four prefectures. The probabilities of crashes and thefts 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 in the form:

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

where y_i is a dummy variable that is defined as either y_i^{Vehicle} or y_i^{Ded}, X_i is a set of attributes of the driver as described in the section "Estimation Results," REGION_{ij} is a set of dummy variables, and ϵ_i is a normally distributed error term. In this equation, we select Iwate as the base group and include dummy variables for each of the remaining

[22] In the case of collision insurance, the estimated coefficient on log (z_i^{Loss}) 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^{Loss}) is −0.07 (0.21) for beginners and 0.08 (0.07) for experienced drivers. Standard errors are in the parentheses.

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

Dependent Variable	y_i^{Vehicle}				y_i^{Ded}			
	Beginners		Experienced Drivers		Beginners		Experienced Drivers	
	Coefficient	p-Value		p-Value		p-Value		p-Value
AGE1	−0.393***	0.000	−0.308***	0.000	−0.196	0.211	−0.190***	0.013
AGE2	0.030	0.624	−0.200***	0.000	−0.432***	0.000	−0.311***	0.000
AGE3	−0.005	0.938	−0.159***	0.000	−0.189*	0.068	−0.186***	0.000
TYPE1	−0.997***	0.004	−0.757***	0.000	−0.342	0.626	−0.247	0.326
TYPE2	−1.467***	0.000	−1.105***	0.000	−0.628	0.363	−0.289	0.244
SIZE1	0.706***	0.014	0.032	0.752	0.158	0.784	0.020	0.918
SIZE2	0.497*	0.083	0.037	0.709	−0.216	0.709	0.051	0.788
SIZE3	0.676**	0.028	0.079	0.460	−0.558	0.354	0.029	0.884
CLASS1	0.362	0.173	0.649***	0.000	1.397***	0.015	0.877***	0.000
CLASS2	0.717***	0.001	1.057***	0.000	0.861**	0.033	0.588***	0.001
CLASS3	0.782***	0.000	1.112***	0.000	0.735*	0.059	0.516***	0.002
CLASS4	0.871***	0.000	1.098***	0.000	0.658*	0.085	0.535***	0.001
CLASS5	0.794***	0.000	1.008***	0.000	0.608	0.114	0.431***	0.008
CLASS6	0.505***	0.016	0.901***	0.000	0.734*	0.069	0.354**	0.037
CLASS7	0.367*	0.091	0.958***	0.000	0.636	0.126	0.267	0.139
CLASS8	0.234	0.310	0.969***	0.000	0.629	0.152	0.414**	0.026
Constant	−0.192*	0.055	0.158***	0.000	0.580***	0.000	1.337***	0.000
REGION1 (Kagoshima)	−0.188	0.133	0.053	0.254	0.397*	0.073	0.218**	0.027
REGION2 (Kanagawa)	−0.042	0.661	−0.004	0.916	−0.218	0.166	−0.204***	0.005
REGION3 (Osaka)	0.028	0.769	−0.057	0.123	−0.183	0.241	−0.256***	0.000
N	2,813		19,184		1,101		10,405	
Log L	−1818.903		−12901.859		−713.619		−3424.086	

Notes: SIZE4 was dropped due to multicolinearity.
***, **, * indicate significance at the 1%, 5%, and 10% level, respectively.

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 in the case of insurance for own vehicle damage (y_i^{Vehicle}). Columns 3 and 4 show that, in the case of deductibles (y_i^{Ded}), drivers in high-risk prefectures are less likely to purchase more coverage. These results are consistent with the results so far, and suggest that uniform pricing across regions does not induce adverse selection.

CONCLUSION

In this article, we set out to investigate whether adverse selection or moral hazard could be induced by the rate regulation that prohibits insurance companies from considering some driver attributes when calculating premiums. Our various empirical studies show no sign of either adverse selection or moral hazard even in the situation that insurance companies are prohibited from considering some driver attributes that seem to a have strong correlation with their risk.[23]

Since the seminal works by Akerlof (1970) and Rothschild and Stiglitz (1976), the insurance market has long been exemplified as one of the markets severely affected by asymmetric information. Later works explained how mechanisms such as risk classification, experience ratings, or deductibles function as devices for mitigating adverse selection or moral hazard. Considering this series of arguments, our empirical results cast some doubt on the role of asymmetric information about individual risk in this market. Interestingly, what is insured, in this market, is not the person but the car.[24] This fact casts further doubt on the role of asymmetric information in this market because it indicates that an insurer does not know who drives the insured car. From these findings, it seems natural to conclude that "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 Salanié, 2000, p. 73).

Two caveats on our results deserve mention. First, although we have carefully checked the robustness of our basic results, the "accident $\neq$ claim" problem is not overcome and the results could be largely affected by this problem. Indeed, the data from the police department imply that this problem is not trivial. According to the police data, the average percentage of cars stolen in the four prefectures is 0.10 percent in 1999 (11,411 cars were stolen out of 11,880,389 cars), whereas our insurance data report that it is 0.22 percent (67 cars were stolen out of 30,000 private-use cars).[25] This implies that insured drivers are more likely to have their cars stolen, which is consistent with the adverse selection story.

Secondly, the results might be specific to the automobile insurance market. Indeed, different empirical results are proposed in other markets. Although Cawley and Philipson (1999) also found evidence against the presence of adverse selection or moral hazard in the life insurance market, there is evidence of these factors in the credit card market (Ausubel, 1999), the annuity market (Finkelstein and Poterba, 2004), and the health insurance market (Barrett and Conlon, 2003). Also, the effects of rate regulation in insurance markets might differ among different markets. For example, Harrington and Danzon (2000) show that rate suppression in workers' compensation insurance brings about an increase in the frequency and/or severity of employee injuries claims, whereas Buchmueller and DiNardo (2002) find that community rating does not induce an adverse selection death spiral in the health insurance market. Therefore, a test for asymmetric information should be carefully conducted in each market.

[23] Although we find no evidence of residual asymmetric information, it does not mean that there is *no* adverse selection or *no* moral hazard in this market. We do not doubt that they exist and matter in some cases: we can easily find articles on insurance frauds in the daily newspaper, for example. What our study indicates is that adverse selection or moral hazard can be managed if insurance companies use some basic mechanisms.

[24] This situation is found not only in Japan but also in other countries.

[25] Figures for private cars are not available.

REFERENCES

Akerlof, G. A., 1970, The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*, 84: 488-500.

Ausubel, L. M., 1999, Adverse Selection in the Credit Card Market, Working Paper, Department of Economics, University of Maryland. Available at: http://www.ausubel.com/creditcard-papers/adverse.pdf/, accessed February, 2006.

Automobile Insurance Rating Organization, 1999, Jidosya Hoken no Gaikyo (An Overview of Japanese Automobile Insurance), in Japanese.

Automobile Insurance Rating Organization, 2002, Automobile Insurance in Japan, (booklet in English).

Barrett, G. F., and R. Conlon, 2003, Adverse Selection and the Decline in Private Health Insurance Coverage in Austria: 1989-95, *The Economic Record*, 79(246): 279-296.

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

Brockett, P. L., and L. L. Golden, 2005, Biological and Psychobehavioral Correlates of Risk Taking, Credit Scores, and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works, Mimeograph.

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*, 92(1): 280-294.

Cawley, J., and T. Philipson, 1999, An Empirical Examination of Information Barriers to Trade in Insurance, *American Economic Review*, 89: 827-846.

Chiappori, P. A., B. Jullien, B. Salanié, and F. Salanié, 2002, Asymmetric Information in Insurance: Some Testable Implications, Mimeograph. Available at: http://www.src.uchicago.edu/pach/cjssdef.pdf/, accessed February, 2006.

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

Chiappori, P. A., and B. Salanié, 2003, Testing Contract Theory: A Survey of Some Recent Works, in: M. Dewatripont, L. P. Hansen, and S. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress* (Cambridge: Cambridge University Press), 114-149.

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

Crocker, K. J., and A. Snow, 1986, The Efficiency Effects of Categorical Discrimination in the Insurance Industry, *Journal of Political Economy*, 94(2): 321-344.

Crocker, K. J., and A. Snow, 2000, The Theory of Risk Classification, in: G. C. Dionne, ed., *Handbook of Insurance* (Boston: Kluwer Academic Publishers), 245-276.

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

Dionne, G. C., 2002, Insurance Regulation in Other Industrial Countries, in: J. D. Cummins, ed. *Deregulating Property-Liability Insurance* (Washington DC: AEI-Brookings Joint Center for Regulatory Studies), 361-396.

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*, 109(2): 444-453.

Finkelstein, A., and J. Poterba, 2004, Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market, *Journal of Political Economy*, 111(2): 183-208.

Harrington, S. E., 1992, Rate Suppression, *Journal of Risk and Insurance*, 56: 185-202.

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

Hayakawa, H., P. S. Fischbeck, and B. Fischhoff, 2002, The Japanese Automobile Insurance Industry—Regulation, Market Structure and Possible Impacts of Deregulation, *Journal of Insurance Regulation*, 18: 385-403.

Hemenway, D., 1990, Propitious Selection, *Quarterly Journal of Economics*, 105: 1063-1069.

Hemenway, D., 1992, Propitious Selection in Insurance, *Journal of Risk and Uncertainty*, 105: 247-251.

Hoy, M., 1982, Categorizing Risks in the Insurance Industry, *Quarterly Journal of Economics*, 97(2): 321-336.

Jaffee, D. M., and T. Russell, 2002, The Causes and Consequences of Rate Regulation in the Auto Insurance Industry, in: J. D. Cummins, ed., *Deregulating Property-Liability Insurance* (Washington DC: AEI-Brookings Joint Center for Regulatory Studies), 361-396.

Koufopoulos, K., 2002, Asymmetric Information, Heterogeneity in Risk Perceptions and Insurance: An Explanation to a Puzzle, *Financial Markets Group Discussion Papers*, dp402.

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*, 104(2): 195-211.

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

Rea, S. A., 1992, Insurance Classifications and Social Welfare, in: G. C. Dionne, ed., *Contributions to Insurance Economics* (Boston: Kluwer Academic Publishers), 377-396.

Rothschild, M., and J. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics*, 90(4): 629-650.

Siegelman, P., 2004, Adverse Selection in Insurance Markets: An Exaggerated Threat, *Yale Law Journal*, 113(6): 1223-1282.

Statistical Bureau of Japan, 1999, Nihon no Tokei (Statistics in Japan), in Japanese.