Testing Adverse Selection with Two-dimensional Information: Evidence from Singapore Auto Insurance Market

Peng Shi[a] Wei Zhang[b†] Emiliano A. Valdez[c]

February 20, 2011

[a] Division of Statistics
Northern Illinois University
DeKalb, IL 60115
Email address: pshi@niu.edu

[b] Department of Economics
Northern Illinois University
DeKalb, Illinois 60115
Email address: wzhang1@niu.edu

[c] Department of Mathematics
University of Connecticut
Storrs, Connecticut 06269-3009
Email address: emiliano.valdez@uconn.edu

[†] Corresponding author
Tel.: +1-815-753-6977
Fax: +1-815-752-1019

Testing Adverse Selection with Two-dimensional Information: Evidence from Singapore Auto Insurance Market

Abstract

This article examines adverse selection in insurance markets within a two-dimensional information framework, where policyholders differ in both their riskiness and degree of risk aversion. Using this setup, we first build a theoretical model to make equilibrium predictions on competitive insurance screening. We study several variations on the pattern of information asymmetry. The outcomes range from full risk separation, to partial separation, to complete pooling of different risk types. Next, we examine results of this construction with an empirical investigation using a cross-sectional observation from a major automobile insurer in Singapore. To test for evidence of adverse selection, we propose a copula regression model to jointly examine the relationship between policyholders' coverage choice and accident occurrence. The association parameter in copula provides evidence of asymmetric information. Furthermore, we invoke the theory to identify subgroups of policyholders for whom one may expect the risk-coverage correlation and adverse selection to arise. The empirical findings are largely consistent with theoretical predictions.

Keywords: Adverse selection, Two-dimensional information, Risk aversion, Insurance screening, Copula models.

1 Introduction

Adverse selection in an insurance market–a stylized situation of asymmetric information–describes the fact that buyers of insurance possess residual private information about their risk that insurers lack even after risk classification. The riskiness of an insurance policyholder is reflected by the probability of a loss, the size of a loss in the event that it occurs, or both. Thus, policyholders of idiosyncratic risks represent different expected indemnity and profitability to insurers, which gives them motivation to try to distinguish or *screen* different types of buyers.

The theory of adverse selection in insurance, set forth by Rothschild and Stiglitz (1976), Stiglitz (1977), and Wilson (1977), derives an intuitively appealing outcome: insurers use deductibles as a screening device to induce customers to reveal their private information, and in equilibrium low risk individuals withstand higher deductibles (less coverage) in order to distinguish themselves from high risk individuals hence to pay a lower premium. The conclusion relies on a condition to align the incentives, typically called the single-crossing property, that is someone with a higher risk has a higher marginal cost in deductibles. Since the seminal work of Rothschild and Stiglitz (1976), the theory of adverse selection has been extended in many significant ways. We refer to the excellent survey articles by Dionne and Doherty (1992) and Dionne et al. (2000) and the literature therein for these contributions.

Much of the empirical literature on adverse selection in insurance markets has been seeking to test the above theoretical prediction on the correlation of risk and coverage: namely, it is the riskier individuals who purchase more insurance coverage. According to Chiappori et al. (2006), the risk-coverage relationship is a quite robust testable implication in a variety of markets. A recent review article by Cohen and Siegelman (2010) summarized the empirical tests on different types of insurance market, including auto insurance, annuities, life insurance, reverse mortgages, health insurance, long-term care, and crop insurance. In particular, the studies of auto insurance market present a mixed testament of the theoretical prediction. Puelz and Snow (1994) found a strong positive risk-coverage correlation, though Dionne et al. (2001) pointed out that the presence of adverse selection vanishes when the nonlinearity of the risk classification variables are taken into account. Using a portfolio of contracts from a French insurer, Chiappori and Salanié (2000) also found no evidence of asymmetric information and the results are robust to both parametric and non-parametric testing methods. The bivariate model proposed by Chiappori and Salanié (1997, 2000) has also been employed by Cohen (2005) and Saito (2006) when analyzing Israel and Japanese auto insurance market, respectively. Supplementing the findings in Chiappori and Salanié (2000), Cohen (2005) concluded with evidence of informational asymmetry for the group of experienced drivers but no coverage-risk correlation for beginning drivers. Saito (2006) found no advantageous information for policyholders when examining the effects of extensive deregulation in the Japanese auto insurance market. Instead of using binary measures, Richaudeau (1999) performed a count regression on number of accidents and found no evidence of adverse selection. In contrast, Kim et al. (2009) showed that a multinomial measure of policyholder's coverage selection could deliver extra evidence on asymmetric information that a dichotomous indicator could not have achieved.

There are several prevailing arguments in explaining the lack of risk-coverage correlation in empirical tests. One of these is that there are factors in insurance purchase decision, noticeably risk aversion, that might offset the positive risk-coverage correlation. Such an explanation needs two necessary assumptions. First, individuals who are more risk averse are inclined to purchase more insurance. Second, risk and risk aversion are correlated in that less risky individuals are also more risk averse. It proceeds to reason that high risk individuals are less likely to purchase insurance because they are also less risk averse. This mechanism might even lead to a negative relation between risk and coverage, which Hemenway (1990, 1992) termed as "propitious selection".

In view of the above argument, we present a theoretical model and subsequent empirical tests using auto insurance data explicitly taking as premise both risk and risk aversion as factors in the decision of insurance purchase. Also motivated by the viewpoint that the existence and degree of adverse selection may vary across insurance or across pools of policyholders within the same insurance (Cohen and Siegelman (2010)), we set forth several variations on the pattern of asymmetric information. In the theory part, we first examine the conventional assumption that buyers of insurance are privately informed of both their individual riskiness and degree of risk aversion. The single crossing condition might be violated under this situation. The equilibrium retains the property of risk separation as in the one-dimensional model if risk aversion matters insignificantly relative to risk in insurance purchase decision. Otherwise, partial pooling of risk types might result. It indicates that the risk-coverage correlation is weaker if risk aversion plays a more pronounced role in insurance purchase, but it does not predict no or negative risk-coverage correlation.

Next, the strong assumption that buyers of insurance must know their exact risk is relaxed. We believe that such an assumption is especially unrealistic in auto insurance, at least for some pools of policyholders. When buyers are only privately informed of their risk aversion, two things become relevant in making equilibrium prediction: 1) the potential correlation of risk and risk aversion, 2) whether individuals are able to infer their expected risk from risk aversion. If the two dimensions of the information space are distributed independently, buyers of insurance have no information advantage regarding their risk, though informed of their risk aversion. In this event, the equilibrium pools risk types. On the other hand, if risk and risk aversion are correlated and individuals are able to make the inference, the equilibrium exhibits the evidence of adverse selection and separation results.

The most pertinent theoretical work to ours is Smart (2000), which similarly treats risk and risk aversion as determinants of insurance purchase decision in a competitive market. What sets our model apart from his is that we also consider a new pattern of asymmetric information in that insurance buyers are only aware of their risk aversion but not their level of risk. We believe that this is a more realistic assumption for many occasions in insurance and fill in a missing gap in existing theory. In another work, Wambach (2000) introduced unobservable wealth in addition to risk as characteristics of policyholders. In addition, de Meza and Webb (2001) and de Donder and Hindriks (2009) added privately known risk aversion into insurance models of moral hazard and asked if propitious selection may result. Jullien et al. (2007) investigated a similar problem in a principal-agent framework. Landsberger and Meilijson (1994) studied a monopolistic insurer's problem when agents differ in their risk aversion. Koufopoulos (2007) derives the conditions under which the positive risk-coverage correlation holds true in a general setting involving multiple loss levels and fixed administrative costs. It is also worth mentioning that multi-period insurance models have been developed for both the competitive and monopolistic markets, and we refer the readers to Dionne et al. (2000) for related research.

Following the existing studies, the evidence of adverse selection in the empirical analysis will be based on the risk-coverage correlation. Extending the widely used bivariate probit regression, we propose a copula regression model to examine the relationship between policyholder's coverage choice and accident occurrence. In addition to the linear relationship, the copula model captures a broader sense of risk-coverage association. Based on a cross-sectional dataset obtained from a major auto insurance company in Singapore, we find a significant positive coverage-risk association, which suggests the existence of adverse selection. Furthermore, we invoke the theory to identify subgroups of policyholders for whom one may expect the risk-coverage correlation and adverse selection to arise. The empirical findings are largely consistent with theoretical predictions: the risk-coverage correlation is significant for experienced drivers but not for beginning drivers; and significant risk-coverage relationship is observed for old drivers but not for young and mid-aged drivers.

Lastly, it is worth mentioning that moral hazard–the *ex post* incentive to exert less effort after purchasing insurance–may lead to the same risk-correlation as adverse selection. Our focus is not to disentangle the effects of adverse selection and moral hazard, and therefore both the theory and empirical analysis shall be considered as the combined effect of both. As pointed out by Abbring et al. (2003), a longitudinal observation is useful in distinguishing adverse selection and moral hazard. Some recent studies along this line include Abbring et al. (2003), Dionne et al. (2005), Israel (2007), and Dionne et al. (2010). Another related group of works on dynamic testing involves learning, where either policyholders or insurers or both learn about policyholder's risk types over time, for example, see Israel (2006) and Cohen (2005, 2008).

The organization of the paper is as follows: Section 2 presents the theoretical model of a competitive insurance market in which buyers are characterized by two dimensions of idiosyncrasy: both their risk and risk aversion. The copula regression model is introduced in Section 3 to jointly examine policyholder's coverage choice and accident occurrence. Section 4 describes the Singapore auto insurance market and presents data characteristics based on a cross-sectional data set obtained from a major auto insurer in Singapore. Section 5 presents the empirical results and economic interpretations. Section 6 concludes the article and discusses possible future work.

2 Economic Theory

2.1 Setup

In this section, we build a theoretical model of insurance market based on Smart (2000), where individuals' idiosyncrasy is two-dimensional: regarding their risk and risk aversion. An insurance market is populated by agents facing idiosyncratic risks of loss and possessing different degrees of risk aversion. Hence agents differ in two dimensions, each assumed to be dual-valued. Individuals, endowed with a wealth $W > 0$, suffer a reduction of amount $D > 0$ with probability π_i in the event of a loss. The loss probability takes two values in the population: $\pi_H > \pi_L > 0$. Each agent has a von Neumann-Morgenstern expected utility on income $u(w, \theta_j)$, which is strictly increasing, strictly concave, and twice-continuously differentiable in its first argument. The degree of risk aversion is indexed by θ_j, where θ_H indicates higher absolute risk aversion than θ_L everywhere. That is, for any $w > 0$

$$-\frac{u_{11}(w,\theta_H)}{u_1(w,\theta_H)} > -\frac{u_{11}(w,\theta_L)}{u_1(w,\theta_L)}$$

Types of individuals are distinguished by their loss probability π_i and degree of risk aversion θ_j, $ij \in T = \{L, H\} \times \{L, H\}$, where T denotes the type space. Hence, an insurance customer can be of low or high risk and of low or high risk aversion. Denote by p_{ij} the percentage of type ij in the population.

All of the above parameters are publicly known to all parties involved in the insurance market. However, insurance companies do not observe individuals' types. Traditionally, the asymmetric information assumption in the insurance literature maintains that insurance customers know privately their types. One point of departure of our model is that we shall also consider the possibility that insurance buyers are only aware of their risk aversion.

Insurance companies offer agents insurance policies that yield W^G in the good state when loss is not incurred and W^B in the bad state when loss takes place. So the induced premium and indemnity of a contract are $W - W^G$, $W^B - W + D$, respectively. We shall impose the requirement of no over-insurance on contracts, that is there is a non-negative deductible in the event of a loss, or $W^G \geq W^B$. Give an insurance contract $C = (W^G, W^B)$, an individual of type ij derives a utility

$$U_{ij}(W^G, W^B) = (1 - \pi_i)u(W^G, \theta_j) + \pi_i u(W^B, \theta_j)$$

In the plane of contracts (W^G, W^B), the indifference curve has a slope of

$$\sigma_{ij}(W^G, W^B) = \frac{dW^B}{dW^G} = -\frac{1-\pi_i}{\pi_i}\frac{u_1(W^G,\theta_j)}{u_1(W^B,\theta_j)}$$

at any point. The slope captures the marginal rate of substitution between the insurance premium and the indemnity. It is easily verifiable that for $ij \in T$, $|\sigma_{Lj}| > |\sigma_{Hj}|$ and $|\sigma_{iL}| > |\sigma_{iH}|$. The first relation is typically described as the single-crossing property of preferences, which says that given the degree of risk aversion, the marginal benefit of the indemnity (or marginal cost of the deductible) is lower everywhere for individuals with lower risk. The second condition is the analogy holding risk constant: the lower the degree of risk aversion, the smaller the marginal benefit of the indemnity. However, while the shapes of indifference curves are readily ranked in each dimension, they are not across dimensions. One observes that there are two possibilities at a given contract C:

$$|\sigma_{LL}| > |\sigma_{LH}| \geq |\sigma_{HL}| > |\sigma_{HH}|$$

or

$$|\sigma_{LL}| > |\sigma_{HL}| > |\sigma_{LH}| > |\sigma_{HH}|$$

where the second case violates the single-crossing property between risks. That is, due to the effect of risk aversion on preferences, it may happen that the marginal cost of deductible is lower for some high risk individuals than some of those who are low risk. This, as anticipated, might render complete separation of risks in equilibrium impossible. Lastly observe that at any full insurance position $W^G = W^B$, the slopes of the indifference curves of both high risk types are the same $-(1 - \pi_H)/\pi_H$, and those for low risk types are $-(1 - \pi_L)/\pi_L$. See Figure 1a for the shape of indifference curves which are labeled by types and point E is the endowment (or no insurance) position.

The insurance industry is assumed to be a perfectly competitive market, where firms are risk neutral and maximize their expected profits:

$$V^i(W^G, W^B) = W - (1 - \pi_i)W^G - \pi_i(W^B + D)$$

from a contract $C = (W^G, W^B)$ offered to an agent with loss probability π_i. So isoprofit curves of contracts offered and accepted by risk π_i types are parallel lines in the contract space with a slope of $-(1 - \pi_i)/\pi_i$. The isoprofit line that crosses the no insurance point represents zero expected profit (or actuarially fair) contracts, and the closer a line is to the origin the higher the profit. See Figure 1b for the shape of fair-odds isoprofit curves, which are labeled by loss probability. The shaded area is the feasible contract space.

2.2 Equilibrium

Following Rothschild and Stiglitz (1976), the interaction between insurance companies and customers is treated as a screening game in which the uninformed firms move first. In the first stage of the game, each firm offers a single insurance contract to the market. In the second stage, customers choose a single policy among all those available. To break ties, assume that when indifferent between different policies, individuals choose the one with higher indemnity and all individuals of the same type choose the same policy if there are multiple identical most preferred policies in the market. We seek the subgame perfect equilibrium of this game.

When types are not known to firms, as motivated before, we investigate two variations on the fashion of asymmetric information: if individuals are fully aware of their two-dimensional types or

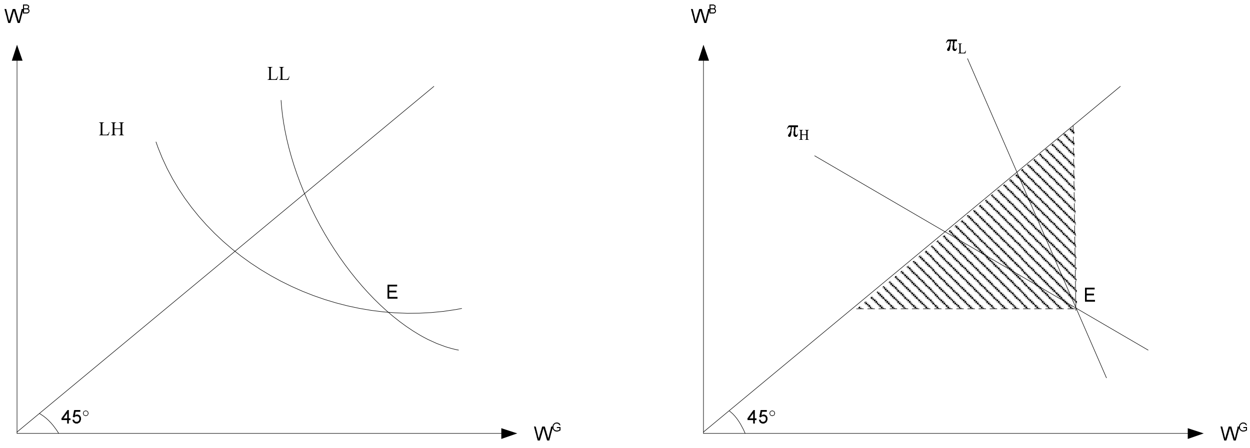

Figure 1: (a) Indifference curves (b) Contract space

if they are only aware of their degrees of risk aversion. The latter is not hard to imagine especially for inexperienced drivers who are yet to understand their risks of driving.

2.2.1 If individuals are fully aware of their two-dimensional types

To begin, it is helpful to state the first best outcome as a reference assuming types are fully observed by insurers. It is well known that the first best requires risk to be completely transferred from a risk averse individual to a risk neutral insurer. Therefore, every type obtains full insurance in equilibrium. Secondly, although different degrees of risk aversion represent different levels of profitability for firms, the force of competition drives away any attempt to differentiate based on risk aversion. In the first best outcome, two actuarially fair full insurance contracts are offered and accepted based on loss probabilities. See Figure 2 where low risk types receive the contract C_L and high risk types receive C_H. To characterize the equilibrium under asymmetric information, we start by presenting some properties the equilibrium outcome must satisfy. For the purpose of empirical tests of this work, we are particularly interested in whether different risks accept different insurance contracts in an equilibrium. If so, an equilibrium shall be referred to as a separating outcome. Otherwise, it is a (partially) pooling outcome.

LEMMA 1. Each insurance contract offered and accepted in an equilibrium must yield non-negative profit. Type LL does not pool with any other type, and type HH does not pool with any low risk type in equilibrium.

PROOF. It is clear that no equilibrium contract shall yield a negative profit because the firm offering it can easily abandon the contract offer and avoid the loss. To show the first half of the second statement, first assume on the contrary that type LL pools with some other type in an equilibrium contract $C = (W^G, W^B)$. Since type LL is the type with the lowest marginal cost of deductible, another offer with larger deductible and smaller premium, $C' = (W^G + \varepsilon_1, W^B - \varepsilon_2)$ where $\varepsilon_1, \varepsilon_2 > 0$, can be made to attract only type LL and be strictly profitable given that the original contract C earns non-negative profit and ε_1 and ε_2 are small enough.

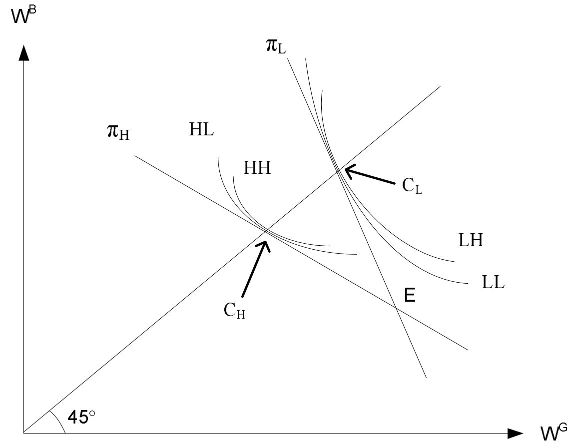

Figure 2: First best contracts

To see that type HH cannot pool with any low risk type, first distinguish two possibilities: type LH and HL accept the same contract or they accept different contracts. In the former case, type HH cannot be in the same pool with type LH and HL because another offer of higher deductible and lower premium can be offered to attract types LH and HL from the pool since type HH has the highest marginal cost of deductible. Such a contract is strictly profitable provided that it is reasonably close to the original contract because a pool of types LH and HL has a lower average loss probability than the original pool of three types. Now consider the case where types LH and HL accept different contracts but type HH pools with type LH in a contract C_{pool}. We first argue that under this assumption, type HL must be receiving its first best contract, i.e. the actuarially fair full insurance offer. Because type HL cannot pool with type LL, it must obtain a contract accepted by no other type. If this is not the first best contract, profitable deviation that attracts type HL exists in the direction of higher premium and lower deductible, regardless of how other types respond. Therefore, type HL must receive its first best contract C_H (see Figure 3). In order for C_{pool} to be incentive compatible for type HL, it must lie below the indifference curve of type HL in Figure 3. However, given the shape of the indifference curve of type HH across C_H, type HH is better off with contract C_H than C_{pool}. This is a contradiction. $Q.E.D.$

Lemma 1 therefore helps in limiting equilibrium possibilities in terms of risk separation to two cases: different risks are separated in an equilibrium, or types LH and HL are pooled so the outcome features partial pooling of risks. Next, each type of equilibrium is characterized.

LEMMA 2. There is a unique separating equilibrium (if it exists). In this equilibrium, high risk types receive their first best contract C_H. Low risk types receive their most preferred insurance contracts, provided that the incentive compatibility constraints of high types are satisfied.

PROOF. It is easy to observe that if risks are separated, high risk types must land on their first best full insurance contract C_H. Because otherwise new strictly profitable offers will be made to drive away any residual insurability or profit. To ensure incentive compatibilities, the contracts low risk types receive must lie below the indifference curves of high risk types in Figure 4a. As in

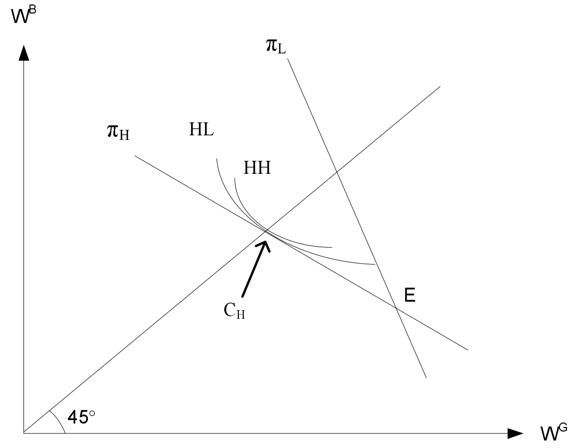

Figure 3: Type HH does not pool with low risk types

the traditional one-dimensional model, this indicates that low risk types are only partially insured in any equilibrium. Also if any incentive compatibility constraint binds in the equilibrium, it has to be type HL's. Next, observe that type HL's incentive compatibility must indeed bind in the equilibrium. This is because that with low risk types only partially insured, a profitable deviation exploring their residual insurability exists, if type HL's incentive compatibility is not yet binding. Following this, type LL must in the equilibrium receive the actuarially fair contract C_{LL}, which is its most preferred offer without violating type HL's incentive compatibility. Suppose instead that it accepts the offer C_{LL}^1. Since type LL has the lowest marginal cost of deductible, another contract C with slightly lower premium and higher deductible can be offered to attract type LL individuals and be profitable. The above can also be said for the contract accepted by type LH in the equilibrium if the single-crossing property is locally satisfied at C_{LL}. If the property is violated at C_{LL}, then C_{LL} cannot be the equilibrium contract for type LH. Because an offer of a higher premium and lower deductible is a profitable deviation that can attract away type LH individuals. Therefore, in the equilibrium type LH must receive C_{LH} (see Figure 4b), which is its most preferred contract and at which type HL's incentive compatibility binds. *Q.E.D.*

It is worth emphasizing that Lemma 2 merely states what a separating equilibrium looks like should it exist. One surprising implication of it is that in a separating equilibrium, the contract offered to type LH yields strictly positive profit when single-crossing is violated in the relevant range, i.e. in Figure 4b contract C_{LH} lies to the left of the zero-profit line denoted by π_L. This occurs in equilibrium because had another contract tried to exploit the residual profitability by attracting away type LH consumers, it would have attracted type HL as well by the violation of single-crossing property at C_{LH}. This would increase the average loss probability hence render the offer unprofitable in some parameter range. [1]

[1] Note that the assumed tie-breaking rule helps in sustaining the equilibrium in the presence of profit in a competitive market. The reason that no other firm would attempt to offer the same profitable contract C_{LH} is that it is assumed that all individuals of type LH choose the same contract so only one such offer survives in the equilibrium.

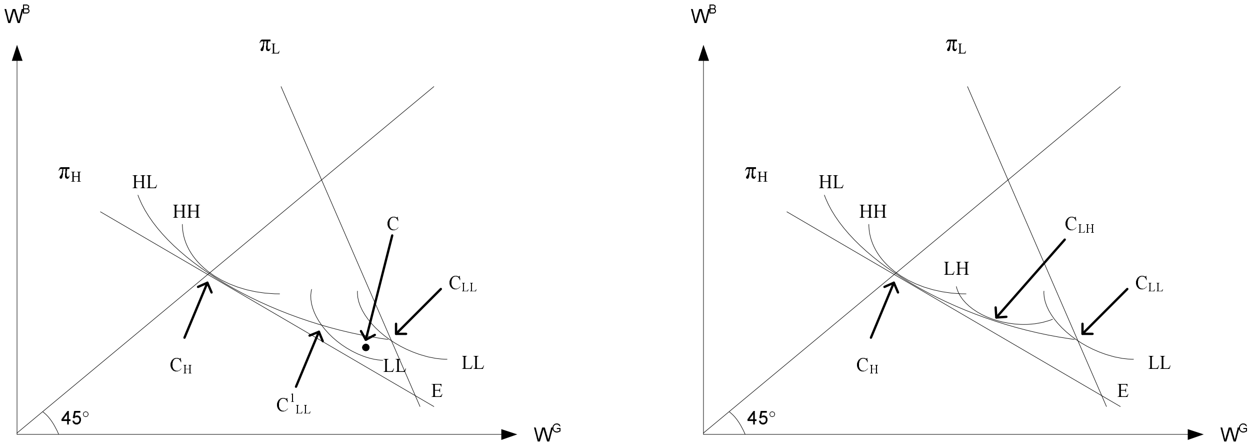

Figure 4: (a) Separating equilibrium (single-crossing satisfied) (b) Separating equilibrium (single-crossing violated)

On the other hand, the argument above also foretells the reason why separation cannot be achieved in other parameter combinations. It occurs in our two-dimensional setup as the force of risk aversion in preferences breaks the single-crossing property which is generally needed to sort out the incentives for proper separation. Define $\bar{\pi} = \frac{p_{LH}}{p_{LH}+p_{HL}}\pi_L + \frac{p_{HL}}{p_{LH}+p_{HL}}\pi_H$ to be the weighted average loss probability of types LH and HL. As shown in Lemma 2, if single-crossing is violated, a separating equilibrium must deliver positive profit on the contract accepted by type LH. However, if C_{LH} is sufficiently away from the isoprofit curve denoted by π_L, in particular if it lies to the left of the isoprofit curve indexed by $\bar{\pi}$ (see Figure 5), separation fails because a contract C_{pool} that attracts both types LH and HL would be a profitable deviation. As implied by Lemma 1, pooling between high and low risks can only happen between types LH and HL, those involved in the violation of single-crossing. The next result characterizes the pooling outcome when separation fails.

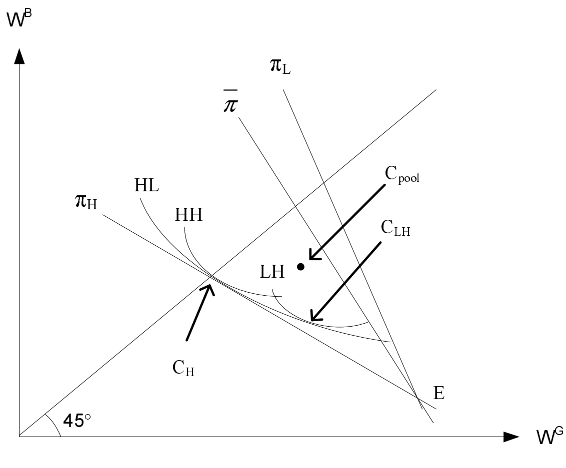

Figure 5: Separation fails

LEMMA 3. There is a unique pooling equilibrium (if it exists). In this equilibrium, type HH receives the first best contract C_H, types LH and HL are pooled in a contract that makes type HH's incentive compatibility binding. And type LL receives a contract that makes type HL indifferent. All contracts yield expected zero profit.

PROOF. The equilibrium is depicted in Figure 6. Recall that $\bar{\pi}$ stands for the average loss probability of types LH and HL. The argument for type HH's contract is identical as in the proof of Lemma 2. Next, types LH and HL must pool in a contract that makes type HH's incentive compatibility bind. Otherwise, there must be a strictly profitable pooling offer to attract away individuals of types LH and HL without affecting type HH's incentive. Also it is worth emphasizing that a prerequisite for risk pooling in equilibrium is the violation of single-crossing condition. If the condition is satisfied locally at C_{pool}, pooling cannot transpire because firms are able to exploit the lower loss probability hence higher profit from type LH by offering a contract with lower premium and higher deductible. This also implies that in equilibrium type LL must receive a contract that makes type HL indifferent. Lastly, as in the case for C_H, both C_{pool} and C_{LL} must yield expected zero profit otherwise the residual profitability must be explored. *Q.E.D.*

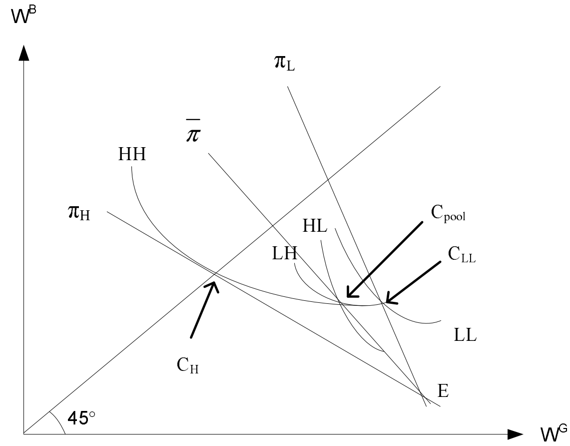

Figure 6: Pooling equilibrium

This result demonstrates an intuitively appealing outcome when single-crossing is violated between risk types: Because of higher degree of risk aversion, type LH is more inclined to acquire insurance than type HL (i.e. type LH has a higher marginal cost in the deductible), therefore they must subsidize type HL in premium in order to obtain an insurance with lower deductible since their preferences render it impossible to acquire lower deductible and separate from type HL simultaneously.

Combining the results of the previous two lemmas, one observes that the *relative* position between the zero profit line indexed by $\bar{\pi}$ and the area where single-crossing is violated plays a key role in determining whether the equilibrium results in separation or pooling of types LH and HL: First, the further right the line is, the more likely the equilibrium is the pooling outcome, *ceteris paribus*. The reason is quite straightforward: as the proportion of type LH individuals increases in the population relative to type HL individuals, the line of $\bar{\pi}$ rotates to the right pivoting against E, the cost to type LH of subsidizing type HL in premium decreases in the pooling equilibrium, hence pooling is more likely to arise. Similarly, fixing everything else, the bigger the area where single-crossing is violated is, or alternatively the more significant risk aversion is in determining preferences, the more likely it is for pooling to occur.

2.2.2 If individuals are only aware of their degrees of risk aversion

The theoretical assumption of asymmetric information–that consumers are privately informed of all relevant information concerning insurance purchase–often proves to be unrealistic in many contexts, noticeably in automobile insurance. Admittedly, knowing the accident probability is a lot to ask from an average driver. Consequently, we also consider another variation on the pattern of asymmetric information: that is, consumers are only aware of their degree of risk aversion. Under this assumption, the population of individuals are grouped by degrees of risk aversion, so there are only two types by private information. We shall use types L and H to describe low and high risk aversion individuals, respectively. Define the population average of loss probability as $\pi = (p_{LL} + p_{LH})\pi_L + (p_{HL} + p_{HH})\pi_H$, and let the average loss probability given risk aversion level θ_j be $\underline{\pi}_j = \frac{p_{Lj}}{p_{Lj}+p_{Hj}}\pi_L + \frac{p_{Hj}}{p_{Lj}+p_{Hj}}\pi_H$. As neither consumers nor insurers have the knowledge of individual loss probability, it becomes relevant whether riskiness can be inferred from risk aversion, that is whether the two dimensions of the type space are correlated. Naturally, there can be the following two possibilities.

Condition 1. The two dimensions of the type space are distributed independently. That is,

$$\frac{p_{LL}}{p_{LL} + p_{HL}} = \frac{p_{LH}}{p_{LH} + p_{HH}}$$

This implies that the two conditional loss probabilities given risk aversion are equal and they are the population average: i.e. $\underline{\pi}_L = \underline{\pi}_H = \pi$.

Condition 2. The two dimensions of the type space are correlated. In particular, a person is more likely to be of low risk given that he is of high risk aversion, and vice versa. That is,

$$\frac{p_{LL}}{p_{LL} + p_{HL}} < \frac{p_{LH}}{p_{LH} + p_{HH}}$$

or consequently, $\underline{\pi}_L > \pi > \underline{\pi}_H$.

Condition 1 treats the two dimensions of the type space as independent variables, hence under this assumption consumers have no information advantage regarding their riskiness although they are privately informed of their risk aversion. Condition 2 concurs with the common notion in the literature that people who are more risk averse also tend to be more cautious in driving thus ultimately incur accidents less frequently. When Condition 2 holds, individuals potentially possess to some extent asymmetric information regarding their riskiness, depending on their understanding of the risk and risk aversion correlation. In what follows, we characterize the equilibrium outcomes under the two assumptions, respectively. In the latter case, we also consider two cases concerning consumers' knowledge of the correlation. We maintain the assumption that insurers, usually with much statistical advantage, comprehend the relation between risk and risk aversion.

First, if condition 1 holds, it is safe to assume that individuals use the population average riskiness in their insurance purchase decision which implies that the slope of the indifference curves of type j is $-\frac{1-\pi}{\pi}\frac{u_1(W^G,\theta_j)}{u_1(W^B,\theta_j)}$. On the other hand, the two subgroups of consumers represent equal profitability to insurers, having equal loss probability. Therefore, with competitive force, the only possible equilibrium is a pooling outcome of all individuals.

LEMMA 4. When Condition 1 holds, all individuals are pooled at a single contract in equilibrium, which is the actuarially fair full insurance.

PROOF. When different degrees of risk aversion implicate the same riskiness, the two types cannot be separated in equilibrium in a competitive market. In addition, the competitive force must drive away any residual profitability or insurability, so the pooling contract has to be the full insurance arrangement which yields zero profit. *Q.E.D.*

Under Condition 2, consumers' beliefs on their individual riskiness is a more delicate issue. Both the arguments that they do or do not have the knowledge of the risk and risk aversion correlation may be justified. Each is examined below in turn. To start, consider the case that individuals are not aware of the correlation and consequently use the population average loss probability in evaluating insurance contracts. One may regard this as a case of asymmetric information in the "reversed" direction because insurers have a more accurate outlook on individual loss probability, knowing the inference of risk aversion, if risk aversion is revealed in equilibrium. Under such circumstance, high risk aversion individuals actually over-estimate their loss probability ($\pi > \underline{\pi}_H$) and those with low risk aversion under-estimate it ($\pi < \underline{\pi}_L$). Because of such biases, it is straightforward to see that in a first-best outcome, insurers would only partially insure low risk aversion individuals (C^L in Figure 7) and over-insure high risk aversion consumers if not subject to any feasibility constraint. Below it is shown that only pooling equilibrium results.

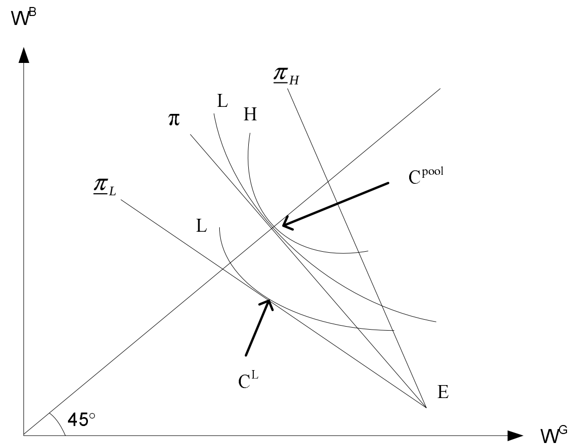

Figure 7: No separating equilibrium and the pooling equilibrium

LEMMA 5. When Condition 2 holds, and consumers use population average riskiness as their individual riskiness, the unique equilibrium is complete pooling, where all individuals receive the actuarially fair full insurance offer.

PROOF. First, we need to show that separation cannot result from the assumptions. By way of contradiction, suppose that there is a separating equilibrium. Then type L individuals, having a high average loss probability, must receive their first best contract C^L (see Figure 7). To ensure incentive compatibility, high risk aversion individuals must receive a contract below the indifference curve of type L crossing C^L. However, a pooling offer close to C^{pool} may attract both types and be strictly profitable. This is a contraction. Next, if the equilibrium must be a pooling outcome, it is the actuarially fair full insurance arrangement. $Q.E.D.$

Lastly, consider the combination of Condition 2 and the assumption that consumers grasp the correlation of risk and risk aversion hence correctly understand the average riskiness indicated by risk aversion. Compared with the usual assumption that individuals privately know both dimensions of their types, the current proposition demonstrates a coarser information advantage on the part of consumers. Although they do not know their exact loss probability, their estimate of it is better than the population average, employing the relation of risk and risk aversion. The slope of the indifference curves of type j changes to $-\frac{1-\underline{\pi}_j}{\underline{\pi}_j}\frac{u_1(W^G,\theta_j)}{u_1(W^B,\theta_j)}$. Notice that it is possible for type H, although with lower average loss probability, to have a higher marginal cost of deductible. In other words, again the single crossing property can be violated. However, surprisingly, the only possible equilibrium is a separating one.

LEMMA 6. When Condition 2 holds, and consumers use the average riskiness implied by the risk aversion level as their individual riskiness, the unique equilibrium (if it exists) is a separating outcome, in which type L receives the first-best full insurance. Type H gets the actuarially fair contract that makes type L's incentive compatibility bind if the single-crossing is locally satisfied, otherwise type H gets their most preferred contract provided that the incentive compatibility of type L is satisfied.

PROOF. First, we need to demonstrate that no pooling is possible. It should be clear by far that any pooling equilibrium of the entire population must yield zero profit. Refer to Figure 8 for potential pooling contracts at actuarially fair rates. At full insurance C_1^{pool}, the slopes of the indifference curves compared to the fair odds as follows

$$|-\frac{1-\underline{\pi}_H}{\underline{\pi}_H}| > |-\frac{1-\pi}{\pi}| > |-\frac{1-\underline{\pi}_L}{\underline{\pi}_L}|$$

therefore the single-crossing holds, which makes pooling impossible. At other places along the fair-odds line, there are two possibilities in terms of how the slopes of the indifference curves compare to the fair odds: either the single-crossing is satisfied, or it is violated so the slopes of both indifference curves are flatter than the fair-odds line. In the former case, pooling cannot result because of the usual argument that type H, with lower average loss probability, has the advantage of a lower marginal cost of deductible. In the latter case (see C_2^{pool}), when the marginal cost of deductible for both types are higher than the fair odds, there is a profitable deviation that offers lower deductible and higher premium that attracts both types. (See Figure 8.)

Next, since the equilibrium (if it exists) must separate the two types, type L therefore, with higher average loss probability, receives its first-best full insurance. Lastly, type H's equilibrium contract must be as specified in the lemma: C_1^H if the single-crossing property is locally satisfied, or C_2^H otherwise (see Figure 9). *Q.E.D.*

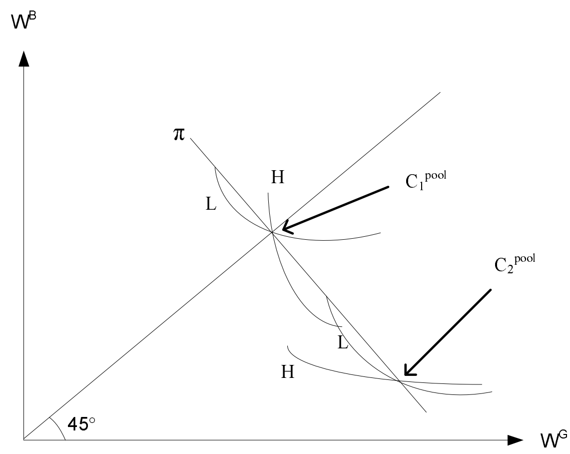

Figure 8: No pooling equilibrium

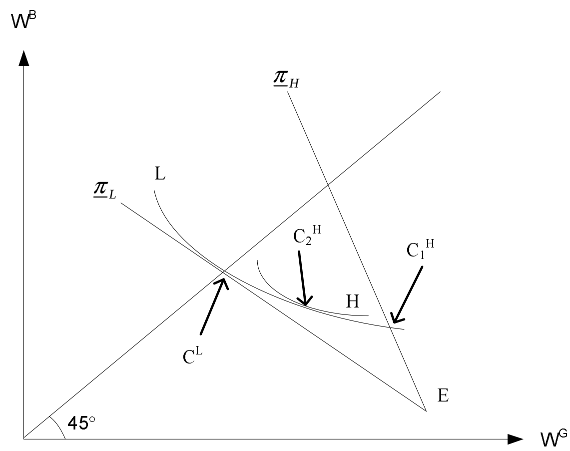

Figure 9: Unique separating equilibrium

3 Econometric Modeling

As stated in the Introduction, our interest is to examine the existence of residual private information for policyholders, where "residual" means after the common practice of risk classification by insurers. The essential idea is to test the independence of coverage choice and accident occurrence, conditional on all the information available to insurers. See Chiappori (2000) and Chiappori and

Salanié (2003) for a broader review on the econometric models and empirical tests related to adverse selection. Different from existing methods, this section presents a copula regression model to test the conditional independence.

Following Chiappori and Salanié (2000), we use a binary variable to indicate a policyholder's coverage and risk. Let y_{i1}^* and y_{i2}^* be the ith insurer's optimal coverage level and inherent risk. These two latent variables are reflected by the policyholder's coverage choice and accident occurrence, denoted by binary variables y_{i1} and y_{i2}, respectively. Here $y_{i1} = 1(0)$ indicates that the policyholder chooses a higher coverage (a lower coverage), and $y_{i2} = 1(0)$ indicates that the policyholder incurs at least one accident (no accident) in the observation period. The relationship between y_i and y_i^* is:

$$y_{i1} = \begin{cases} 0 & y_{i1}^* \leq 0 \\ 1 & y_{i1}^* > 0 \end{cases} \text{ and } y_{i2} = \begin{cases} 0 & y_{i2}^* \leq 0 \\ 1 & y_{i2}^* > 0 \end{cases}. \tag{1}$$

Assuming a linear relationship for both latent variables y_{i1}^* and y_{i2}^*, we have the following expressions:

$$\begin{aligned} y_{i1}^* &= \boldsymbol{x'\beta} + \varepsilon_{i1} \\ y_{i2}^* &= \boldsymbol{z'\gamma} + \varepsilon_{i2} \end{aligned},$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ represent the vector of explanatory variables that affect the coverage choice and risk level of the policyholder, respectively. The terms $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are coefficients to be estimated. The error terms ε_{i1} and ε_{i2} are assumed to be independent with explanatory variables. Thus, the joint mass probability function of y_{i1} and y_{i2} is:

$$\begin{aligned} f_i(0,0) &= \text{Prob}(y_{i1} = 0, y_{i2} = 0|\boldsymbol{x}, \boldsymbol{z}) = \text{Prob}(\varepsilon_{i1} \leq -\boldsymbol{x'\beta}, \varepsilon_{i2} \leq -\boldsymbol{z'\gamma}) \\ f_i(1,0) &= \text{Prob}(y_{i1} = 1, y_{i2} = 0|\boldsymbol{x}, \boldsymbol{z}) = \text{Prob}(\varepsilon_{i1} > -\boldsymbol{x'\beta}, \varepsilon_{i2} \leq -\boldsymbol{z'\gamma}) \\ f_i(0,1) &= \text{Prob}(y_{i1} = 0, y_{i2} = 1|\boldsymbol{x}, \boldsymbol{z}) = \text{Prob}(\varepsilon_{i1} \leq -\boldsymbol{x'\beta}, \varepsilon_{i2} > -\boldsymbol{z'\gamma}) \\ f_i(1,1) &= \text{Prob}(y_{i1} = 1, y_{i2} = 1|\boldsymbol{x}, \boldsymbol{z}) = \text{Prob}(\varepsilon_{i1} > -\boldsymbol{x'\beta}, \varepsilon_{i2} > -\boldsymbol{z'\gamma}) \end{aligned}$$

Since a policyholder's coverage selection and risk type are endogenously determined, it will be more efficient to model the two equations simultaneously. One can specify a bivariate probit model by assuming a joint normal distribution for ε_{i1} and ε_{i2}. Then the relationship between y_{i1}^* and y_{i2}^* is measured by the correlation coefficient in the bivariate normal distribution. For example, see Chiappori and Salanié (2000) and Saito (2006). However, the Pearson correlation only captures a linear relationship. The absence of linear correlation is not a sufficient condition for conditional independence. To relax the bivariate normal assumption and allow for a broader concept of association between y_{i1}^* and y_{i2}^*, we propose using a parametric copula function to jointly model policyholder's coverage choice and accident occurrence.

A copula is a multivariate distribution with all marginals following uniform distribution on [0,1]. Copulas are a useful tool in the multivariate analysis and they accommodate both linear and non-linear relationships (see Nelsen (2006) and Joe (1997) for comprehensive reviews on copulas). Copula regression begins in the social science literature with Frees and Wang (2005, 2006). In insurance, the most recent work on copula regression has been on predictive modeling, for example, see Frees and Valdez (2008), Frees et al. (2009), and Shi and Frees (2010). In spite of the aforementioned work, the application for discrete data is scanty. Some recent studies within social science contexts include Prieger (2002), Smith (2003), Cameron et al. (2004), Purcaru and Denuit (2005), and Zimmer and Trivedi (2006).

By way of a parametric copula, denoted by $C(\cdot;\theta)$ with parameter θ, the joint mass probability function of y_{i1} and y_{i2} can be expressed as:

$$\begin{aligned} f_i(0,0) &= C[G_1(-\boldsymbol{x'\beta}), G_2(-\boldsymbol{z'\gamma})] \\ f_i(1,0) &= C[1, G_2(-\boldsymbol{z'\gamma})] - C[G_1(-\boldsymbol{x'\beta}), G_2(-\boldsymbol{z'\gamma})] \\ f_i(0,1) &= C[G_1(-\boldsymbol{x'\beta}), 1)] - C[G_1(-\boldsymbol{x'\beta}), G_2(-\boldsymbol{z'\gamma})] \\ f_i(1,1) &= 1 - C[1, G_2(-\boldsymbol{z'\gamma})] - C[G_1(-\boldsymbol{x'\beta}), 1] + C[G_1(-\boldsymbol{x'\beta}), G_2(-\boldsymbol{z'\gamma})] \end{aligned} . \tag{2}$$

Here, G_1 and G_2 are the cumulative distribution functions of ε_{i1} and ε_{i2}, respectively. Two commonly used forms of G for a binary specification are probit $G(a) = \Phi(a)$ and logit $G(a) = 1/(1+\exp(-a))$. It is straightforward to show that model (2) reduces to a bivariate probit model when $G(a) = \Phi(a)$ and $C(\cdot)$ is a Gaussian copula.

The model by its nature is fully parametric and can be easily estimated using a likelihood-based estimation method. The log likelihood function can be derived by summing up the logarithm of joint probability function (2) over all policyholders. The theory of adverse selection predicts a non-negative relationship between coverage choice and accident occurrence. To examine whether the empirical result is consistent with the theoretical prediction, one cannot restrict to copulas that only allow for positive association. Motivated by such fact, we consider the Frank copula that renders more flexibility in the dependence modeling:

$$C(u_1, u_2) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta\mu_1} - 1)(e^{-\theta\mu_2} - 1)}{e^{-\theta - 1}}\right\},$$

where θ is the dependence parameter that captures the association between two responses. When $\theta \to 0$, the Frank copula becomes the product copula, indicating the independence case. $\theta > 0$ ($\theta < 0$) indicates that the two responses are positively (negatively) associated. As a robust test in the empirical analysis, we also examine the conditional independence using other forms of copulas.

4 Data

4.1 Singapore Auto Insurance

In Singapore, auto insurance represents one of the largest class of business underwritten by general insurers. According to the General Insurance Association of Singapore (GIA), the industry representative body of all non-life insurance companies, the gross premium income for auto insurance in

2009 is $0.8 billion (Singapore dollars), taking up a 36 percent share of the entire market.

Just as in many parts of the world, some form of auto insurance is mandatory for anyone who drives and owns a car in Singapore. In general, what is mandatory is a minimum level of 'third party' insurance which covers death or bodily injury to third parties. Typically, 'third party' coverage provides additional protection against liability that may arise as a result of damage to third party property. Apart from third party benefits, fire and theft may also provide additional coverage against damage from fire or theft. As is often the case, car owners also buy a comprehensive policy which additionally covers damage to the insured vehicle, and in many cases, medical expenses, in the event an accident occurs.

A very important feature of an auto insurance policy in Singapore is the 'no claims discount' (NCD) scheme. Introduced in 1982, the scheme was designed to encourage good driving behavior. The discount is applied to the premium level of what is considered a standard policy, in increments of 10%. Each year the policy makes no claim, this 10% discount is applied, subject to a maximum of 50%. This reduction is usually dramatically reduced in the event a claim or multiple claims are made. The NCD scheme is quite similar in principle to the bonus-malus scheme typically found in auto insurance policies in Europe.

4.2 Variable Description

We consider a cross-sectional observation of policyholders from a major automobile insurance company in Singapore. Our sample consists of a total of 15,418 policies that were in effect in year 2001. To test the coverage-risk relationship, we differentiate a policyholder's coverage choice according to whether or not a comprehensive policy is purchased, and a policyholder's riskiness as to whether or not an accident occurs in the observation period. Among the 15,481 policyholders, 74.38 percent (11,468 contracts) purchased comprehensive coverage, and the remaining 25.62 percent (3,950 contracts) purchased third party coverage. In terms of accident occurrence, the majority of policyholders, i.e. 13,621 drivers accounting for 88.34 percent of observations, did not have any accident during the year, and the rest 11.66 percent of policyholders (1,797 contracts) incurred at least one accident.

To obtain an initial idea of the coverage-risk correlation, we display a two-way frequency table of policyholder's coverage choice and accident occurrence in Table 1. On the one side, of the policyholders with comprehensive coverage (high coverage), 13.72 percent incurred one or more accidents, while of the policyholders with third party coverage (lower coverage), 5.67 percent incurred one or more accidents. On the other side, among the policyholders with accidents (high risk), 87.53 percent buy comprehensive policy, while among the policyholders without accidents (low risk), 72.65 percent buy comprehensive coverage. This observation is consistent with the χ^2 statistic and ϕ coefficient, which suggests a positive relationship between coverage choice and accident occurrence. Note that this relationship does not provide any evidence of asymmetric information. The positive correlation could be caused by some other exogenous variables, i.e., the information to available to both insurers and insurees. To show the presence of asymmetric information, one needs to look into the residual relationship after controlling for the exogenous variables.

Table 1: Frequency table for coverage choice and accident occurrence

		Accident Occurrence				
		0	1	Total	χ^2 statistic	184.69
	0	3726	224	3950	p-value	<0.0001
Coverage Choice	1	9895	1573	11468	ϕ coefficient	0.1095
	Total	13621	1797	15418		
		Beginning Drivers				
		Accident Occurrence				
		0	1	Total	χ^2 statistic	19.74
	0	437	30	467	p-value	<0.0001
Coverage Choice	1	881	149	1030	ϕ coefficient	0.1148
	Total	1318	179	1497		
		Experienced Drivers				
		Accident Occurrence				
		0	1	Total	χ^2 statistic	165.68
	0	3289	194	3483	p-value	<0.0001
Coverage Choice	1	9014	1424	10438	ϕ coefficient	0.1091
	Total	12303	1618	13921		

In addition, we classify the policyholders into beginning drivers and experienced drivers. Following Chiappori and Salanié (2000) and Cohen (2005), the beginning drivers are those with less than three years of driving experience, and the experienced drivers are those with more than three years of driving experience. At least three reasons for differentiating these two types are discussed in the literature : First, beginning drivers may know less about their true riskiness and driving skills that could affect the occurrence of accidents; Second, experienced drives are more knowledgable about the coverage of the insurance policy for their vehicles; Third, driving experience is presumably related to the age of the policyholder. A larger possibility of homogeneity could be achieved in a population of less different seniority groups. The two-way frequency tables for these two groups are also provided in Table 1. Similar to the result in the first panel, we observe significant positive association between coverage choice and accident occurrence for both beginning and experienced drivers, though the χ^2 statistic of the former is smaller. The higher ϕ coefficient for experienced drivers foreshadows the results found in Cohen (2005), i.e. the selection effect is weaker for beginning drivers. The formal test of this hypothesis will be discussed in the following section.

Because the data used to study the behavior of drivers are from an insurance company, it is worth stressing the distinction between an accident and a claim. An accident will not become a claim if not reported to the insurer. Whether or not a policyholder will report an accident is the policyholder's discretion. For example, a policyholder typically does not report an accident when the amount of damage is smaller than the deductible. With regard to accident occurrence, we consider reported accidents rather than actual accidents, since the occurred but not reported accidents are not available from the insurer. We note that some researchers limit the observations to "bilateral accidents" to mitigate the bias caused by the potential underreporting. This practice is not necessarily appropriate for our data set: First, the coverage choice in our analysis is measured by thirty part vs. comprehensive policy rather than high deductible vs. low deductible. Discarding the "unilateral" accidents that are covered by comprehensive policies could underestimate the risk of these policyholders. Second, existing evidence shows that the conditional correlations do not differ statistically whether the occurrence is measured by reported accidents or actual accidents (for example, see Richaudeau (1999)). Third, from the perspective of insurers, the claim might be the more appropriate measure of the risk of policyholders, because it is the claims that ultimately determine the expected payout of insurers. Finally, the empirical evidence presented below regarding subgroups of policyholders reinforces the assumption that the bias due to the discrepancy between accidents and claims is immaterial.

In addition to the coverage choice and accident history, our data provide an extensive set of information for each policy, including the policyholder's demographic characteristics (age, gender, marital status), the insured vehicle's characteristics (engine size, main usage, brand), and the policyholder's driving characteristics (years of driving, NCD). The rich data set allows us to control many potential exogenous variables in detecting the residual adverse selection in various risk classes. The definitions and descriptive statistics of both endogenous and exogenous variables are displayed in Table 2. According to these descriptive statistics, we may conclude that a typical observation in our data would be one who is middle aged (between 36 and 45) and a male driver who is married with roughly 12 years of driving experience. The typical vehicle driven in our portfolio is roughly 7 years old, a private car with medium capacity whose brand is either a Toyota or something else other than those listed in the table. To motivate our model specification, we decompose the data by coverage choice and risk level of policyholders, and present the summary statistics for each category in Table 2 as well. We observe significant difference in both driver characteristics and vehicle characteristics across different groups of policyholders. For example, those policyholders who choose third party on average have lower NCD while those who choose comprehensive on average have higher NCD.

An important exogenous variable which needs to be controlled for when estimating the conditional coverage-risk correlation is the insuree's past driving records. The driving history contains useful information on the accident probability and is typically employed by insurers in the determination of premiums for the contract (e.g., see Boyer and Dionne (1989) and Dionne and Vanasse (1992)). As pointed out by Chiappori and Salanié (2000), when such information is available to both insurers and insurees, failing to control for the insuree's past driving history could overestimate the level of asymmetric information. For example, the insured's driving records were omitted in Puelz and Snow (1994), suggesting that the observed positive coverage-risk relationship could be explained by the fact that public information of risk type is mistakenly treated as private. One commonly used proxy for past driving records in recent studies is the bonus-malus coefficient (for example, see Saito (2006), Li et al. (2007), and Kim et al. (2009)). The bonus-malus system is popular in European automobile insurance markets. Some Asian markets also use bonus-malus system, such as Japan and South Korea. In essence, the accident history is incorporated in the driver's bonus-malus coefficient, which is then applied to a basic amount to determine the premium for the next period. The way the coefficient is determined makes the bonus-malus rate a good indicator of the insuree's accident history for the past several years. In the empirical analysis, we control for the insuree's past driving history using the NCD. As introduced in Section 4.1, the NCD in Singapore auto insurance market acts in a similar vein as the bonus-malus coefficient in European markets, encouraging safe driving.

Table 2: Descriptive statistics of response and explanatory variables

Dependent Variables		Mean	StdDev								
Coverage	=1 if comprehensive coverage, 0 if third party	74.38%									
Risk	=1 if accident occurs, 0 if no accident	11.66%									
Independent Variables				Third Part		Comprehensive		No Accident		≥ 1 Accident	
				Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
Ageinsured	the age of the policyholder	41.99	10.39	42.56	11.23	41.80	10.08	42.20	10.40	40.44	10.24
Sexinsured	=1 if the policyholder is female, 0 if male	17.87%		13.01		19.54		17.84		18.09	
Marital	=1 if the policyholder is married, 0 if single	84.92%		84.61		85.03		85.18		82.92	
Vage	the age of the insured vehicle	7.13	6.03	15.44	5.07	4.27	2.85	7.33	6.15	5.63	4.75
vehicleclass	=1 if the vehicle is a private car	84.86%		74.91		88.29		84.74		85.81	
	=2 if the vehicle is a goods vehicle	14.15%		22.91		11.13		14.17		13.97	
	=3 if others (reference level)										
capacityclass	=1 if petty cars (cubic capacity is less than 1000 or tonnage is less than 1)	11.68%		18.94		9.18		12.08		8.68	
	=2 if small cars (cubic capacity is between 1000 and 1500 or tonnage is between 1 and 2)	33.34%		31.49		33.97		33.13		34.89	
	=3 if medium cars (cubic capacity is between 1500 and 2000 or tonnage is between 2 and 3)	47.76%		41.65		49.87		47.55		49.36	
	=4 if large cars (cubic capacity is greater than 2000 or tonnage is greater than 3) (reference level)										
brandclass	=1 if Toyota	20.03%		26.1		17.94		20.38		17.36	
	=2 if Honda	14.63%		23.47		11.59		14.51		15.58	
	=3 if Nissan	17.54%		13.95		18.78		17.56		17.42	
	=4 if Misubishi	9.65%		6.58		10.71		9.49		10.85	
	=5 if Mazada	4.57%		3.44		4.96		4.54		4.84	
	=6 if other Japanese car	3.79%		4.2		3.65		3.94		2.73	
	=7 if Korean car	6.60%		0.61		8.67		6.32		8.74	
	=8 If European car	19.30%		19.37		19.27		19.37		18.75	
	=9 if others (reference level)										
experience	length of driving experience of the insured	12.67	8.67	12.07	8.82	12.88	8.61	12.76	8.69	12.00	8.51
NCD	No claims discount enjoyed by the insured										
	=1 if 0 percent	24.70%		38.51		19.94		24.47		26.43	
	=2 if 10 percent	14.34%		16.99		13.43		14.06		16.47	
	=3 if 20 percent	11.69%		12		11.59		11.78		11.07	
	=4 if 30 percent	7.87%		5.72		8.62		7.77		8.63	
	=5 if 40 percent	6.78%		4.13		7.7		6.64		7.85	
	=6 if 50 percent (reference level)										

5 Empirical Results

We first test for the presence of private information of policyholders using the pool of all observations from the insurer. Model (2) is estimated with different specifications. Specifically, we consider probit and logit for both marginals, and based on the Akaike's Information Criterion (AIC) statistic, we choose the model with the best fit as shown in Table 3. Note that in the final model, a logit regression is more appropriate for policy choice and a probit regression is more appropriate for accident occurrence. This result also emphasizes the flexibility of using copulas in the specification of marginal choice model.

Table 3: Estimation results for the copula model

	Coverage - Logit			Risk - Probit		
	Estimate	StdErr		Estimate	StdErr	
INT	6.8538	0.5101	***	-1.5400	0.2400	***
AGEINSURED	-0.0032	0.0055		-0.0050	0.0017	***
SEXINSUREDF	0.1795	0.1298		-0.0264	0.0354	
MARITALM	0.1426	0.1332		-0.0155	0.0398	
VAGE	-1.1471	0.0306	***	-0.0297	0.0027	***
VEHICLECLASS1	5.0233	0.3518	***	0.6757	0.2168	***
VEHICLECLASS2	3.0191	0.3678	***	0.6694	0.2189	***
CAPACITYCLASS1	-0.7095	0.2551	***	-0.1136	0.0703	
CAPACITYCLASS2	0.2171	0.2165		-0.0086	0.0573	
CAPACITYCLASS3	0.3840	0.2178	*	0.0209	0.0544	
BRANDCLASS1	0.6266	0.2197	***	-0.0403	0.0771	
BRANDCLASS2	0.2224	0.2229		0.1044	0.0779	
BRANDCLASS3	0.8661	0.2339	***	-0.0210	0.0767	
BRANDCLASS4	0.6201	0.2518	**	0.0113	0.0816	
BRANDCLASS5	0.5039	0.2956	*	-0.0192	0.0934	
BRANDCLASS6	0.8793	0.2782	***	-0.1309	0.1051	
BRANDCLASS7	0.9710	0.3209	***	0.0678	0.0853	
BRANDCLASS8	1.0297	0.2324	***	0.0353	0.0776	
EXPERIENCE	-0.0098	0.0061		-0.0010	0.0019	
NCD0	-1.2394	0.1332	***	0.1666	0.0411	***
NCD10	-0.9393	0.1530	***	0.1864	0.0448	***
NCD20	-0.6116	0.1613	***	0.0827	0.0481	*
NCD30	-0.1680	0.1888		0.1352	0.0526	**
NCD40	-0.2419	0.2081		0.1560	0.0550	***
Dependence	2.3243	0.3791				
Log-likelihood	-7050.98					
AIC	14199.96					
χ^2-statistic	44.042					

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

Tables 3 shows that many exogenous variables have a statistically significant effect on either policyholder's coverage choice or accident occurrence. A special note is needed for some exogenous variables. First, we examine policyholder's choice of insurance coverage. The non-significant effect of insuree's age could be explained by the combined results of risk-effect and price-effect (see Richaudeau (1999)). Young drivers presumably treat themselves as high risk and thus intend to purchase comprehensive policy, since they do not have enough driving experience. Meanwhile, representing higher risk, young drivers will be charged higher premiums, which reduces the demand for comprehensive coverage. The negative coefficient, though not statistically significant, suggests the dominance of the risk-effect. Another factor that might influence the relationship between age and coverage choice is the driver's risk attitude. Other things equal, the more risk averse the policyholder is, the more coverage he will purchase. The insuree's age is an important indicator of the degree of risk aversion, and thus could introduce a non-linear effect of age on coverage choice. Although the effects are not statistically significant, the positive sign for the coefficients of sex and marital status suggests that female and married policyholders prefer a more comprehensive coverage. This could be explained by the higher degrees of risk aversion of the two groups. The age of the vehicle exhibits significant effect on the policy choice. As the vehicle depreciates over time, one tends to purchase less coverage for it. As expected, the vehicle's characteristics (type, capacity, brand) play important roles in the policyholder's decision on coverage choice. Consistent with the risk-effect, we observe a negative, though not statistically significant coefficient of the policyholder's driving experience. The NCD also appears to be an important determinant in the decision of insurance purchase. The significant effect of NCD (NCD=0, 10, 20) implies that a policyholder with a high NCD would buy more insurance for the vehicle. This is because for a higher discount makes the same policy cheaper.

Next we examine the occurrence of accidents. Not surprisingly, the effect of insuree's age is statistically significant. Typically, young drivers have fewer years of driving experience and are believed to have the tendency to drive less carefully, and thus are considered high risks drivers. The negative sign of the coefficients of sex and marital status suggests that female and married drivers have smaller chance to incur accidents. However, as in the case of coverage choice, the effects of both variables are not statistically significant. Perhaps counterintuitively, we observe a negative effect of vehicle's age on the probability of accident. This negative relationship simply suggests that safe drivers generally drive older cars, according to our data. As for the vehicle's characteristics, we observe a significant effect of the type, but not for the size and brand. This is not surprising, one does not expect to see that the brand of the automobile affects the probability of accidents, because the propensity of accident relies more on the driver's behavior. The coefficient of the driver's experience is negative but not statistically significant. The negative sign is expected since more experienced drivers should have better skill and thus lower accident probability. The NCD is an indicator of a policyholder's accident history. Consistently, a driver with a lower NCD (compared with NCD=50) has a higher chance of accidents. Note that we only discuss the effect of NCD compared with the reference level.

To test the presence of asymmetric information, we examine the association between policyholder's coverage choice and accident occurrence. A positive correlation indicates the existence of adverse selection. First and foremost is the examination of the association parameter θ in the Frank copula. Recall that a positive (negative) θ indicates a positive (negative) relationship. A simple t statistic could be used to show the statistical significance of this association. Our estimation produced an estimated dependence parameter of 2.32 in the Frank copula, which translates to a Spearmans ρ of 0.36. A corresponding standard error of 0.38 suggests a statistically significant positive association between coverage choice and accident occurrence. The second test that we perform is a likelihood ratio test. Since the Frank copula nests the product copula as a special case, we re-estimate the bivariate copula model with the product copula instead of the Frank copula. The χ^2 statistics is 44.04, leading to a consistent result with the t test. In conclusion, both statistical tests imply the presence of advantageous information for policyholders.

A final examination of the copula model is a robustness test. The above knowledge about the coverage-risk relationship suggests fitting a model based on the Frank copula which allows for both positive and negative association. Due to the non-linear optimization required in the evaluation of the likelihood function, we are interested in the robustness of the coverage-risk correlation by examining other copula specifications. In doing so, we re-calibrated the copula model under two other customarily used Archimedean-type copulas, the Gumbel copula and the Clayton copula with the following respective specifications:

$$\text{Gumbel: } C(u_1, u_2; \theta) = \exp\left[-((-\log u_1)^\theta + (-\log u_2)^\theta)^{1/\theta}\right], \ \theta \geq 1$$

$$\text{Clayton: } C(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \ \theta > 0$$

where θ represents the dependence parameter in both copulas. Note that unlike the Frank copula, the Gumbel and Clayton copulas do not accommodate negative association. The estimated positive relationship that we have already observed based on the Frank copula between policy choice and risk suggests that both copulas are eligible to test for possible robustness. It is easy to see that we have the case of independence when $\theta = 1$ for the Gumbel copula, and when $\theta \to 0$ for the Clayton copula.

Since our main interest is to test the significance of the coverage-risk correlation, we only report the result of t and χ^2 statistics in Table 4. Unlike the Frank copula, the statistical tests using both Clayton and Gumbel copulas involve boundary hypothesis. Therefore defining the t-statistics as the estimates minus the boundary and divided by the standard error, we perform a one-side t test for both Clayton and Gumbel copulas. Though the asymptotic distribution in the likelihood ratio tests does not strictly follow Chi-square distribution, the large statistics suggest the estimation bias is negligible. All the findings support that the positive coverage-risk association is quite robust to the specification of copulas.

Table 4: Robust test on the coverage-risk relationship

	t test					
	Frank		Clayton		Gumbel	
	t-stat	p-value	t-stat	p-value	t-stat	p-value
All drivers	6.13	<0.01	4.05	<0.01	4.99	<0.01
Beginning	1.42	>0.10	1.34	>0.05	1.28	>0.05
Experienced	5.92	<0.01	3.81	<0.01	4.81	<0.01
Young	1.54	>0.10	0.76	>0.10	1.50	>0.05
Mid-age	5.62	<0.01	3.99	<0.01	4.56	<0.01
Old	1.81	>0.05	1.12	>0.10	1.51	>0.05
	χ^2 test					
	Frank		Clayton		Gumbel	
	χ^2-stat	p-value	χ^2-stat	p-value	χ^2-stat	p-value
All drivers	44.02	<0.01	27.60	<0.01	41.77	<0.01
Beginning	1.94	>0.10	1.03	>0.10	3.31	>0.05
Experienced	41.44	<0.01	24.39	<0.01	39.96	<0.01
Young	2.01	>0.10	0.12	>0.10	3.81	>0.05
Mid-age	37.41	<0.01	30.61	<0.01	32.75	<0.01
Old	3.02	>0.05	1.26	>0.10	3.24	>0.05

5.1 Drivers by Experience

Informed by the theoretical results developed in Section 2, we seek pools of policyholders for whom it may be more appropriate to believe that both risk and risk aversion are their private knowledge, and for whom it may be suitable to assume that knowing their riskiness is beyond their reach. We think that driving experience serves as a good indicator of one's knowledge of the risk of driving.

As noted earlier, Chiappori and Salanié (2000) performed a test using bivariate probit model for policyholders with less than three years of driving experience, and found no coverage-accident correlation for the group of policyholders. Similarly, Cohen (2005) examined the correlation for both beginning and experienced drivers, and found consistent results with Chiappori and Salanié (2000), i.e., a non-significant correlation for beginning drivers and significant correlation for experienced drivers. In the same line, we split policyholders into two sub-groups, according to whether a policyholder has less than three years of driving experience. We end up with a relatively small sample for beginning drivers and a large sample for experienced drivers. The Frank copula model is estimated for both groups and the results are displayed in Table 5. Consistent with the above two studies, we find a significant risk-coverage correlation for experienced drivers, but not for beginning drivers. We attribute these observations to the fact that beginning drivers only know their degrees of risk aversion but have little knowledge about their true level of riskiness. In contrast, experience drivers have more information about their risk type as well as their risk attitude. Such findings are firmly supported by the equilibrium predictions of our theoretical model: when agents are only privately informed of their risk aversion and cannot infer expected risk from risk aversion, the equilibrium pools the risk types; and when agents are privately informed both of their risk and risk aversion and risk aversion matters insignificantly to risk, the equilibrium separates risk types. A robust test is also performed using other forms of copulas and the corresponding statistical tests are displayed in Table 4. Under both Clayton and Gumbel copulas, beginning drivers are pooled together and experience drivers are separated according to their risk types.

5.2 Drivers by Age

When seeking pools of policyholders with varying strength of the role of risk aversion in insurance decision, we use age as the indicator. Because it is commonly believed as one gets older, risk aversion plays an increasingly significant role in insurance purchase and other decisions involving uncertainties (see for example Morin and Suarez (1983) and Pålsson (1996)).

To provide evidence of the risk aversion effects, we divide the data into three subsets according to insured's age: young (less than 25), mid-age (between 25 and 55), and old (greater than 55). The copula model is calibrated using the three subsets individually and we examine the association parameter of the Frank copula for each group. Recall that the theoretical model predicts a separating equilibrium when risk aversion matters insignificantly and a partial pooling otherwise. Presumably age is a good indicator of the degree of risk aversion, we expect to see a weaker coverage-risk correlation for the older group. Table 6 summarizes the estimation results of the Frank copula model for the three age groups as well as the associated likelihood ratio test. The null hypothesis is that there is no relationship between the choice of coverage and the occurrence of accidents. As we see, this null hypothesis cannot be rejected for both young and old groups at least at 5% significant level. However, the independence hypothesis is firmly rejected for the mid-age group. This result is supported by the theory of risk aversion effect: because of the higher degree of risk aversion in the old group, one expects to see a greater offsetting effect. The insignificant correlation for the young drivers is explained by the fact that most young drivers are also beginning drivers. The robust test is provided in Table 4, and again our findings are not sensitive to copula specifications.

Table 5: Estimation of the copula model by driving experience

Beginning Driver			Experienced Driver		
	Estimate	StdErr		Estimate	StdErr
CHOICE_INT	10.2630	2.2440	CHOICE_INT	6.5694	0.5441
CHOICE_AGEINSURED	0.0114	0.0188	CHOICE_AGEINSURED	-0.0110	0.0067
CHOICE_SEXINSUREDF	0.2430	0.4000	CHOICE_SEXINSUREDF	0.1746	0.1393
CHOICE_MARITALM	-0.0049	1.0454	CHOICE_MARITALM	0.1303	0.1422
CHOICE_VAGE	-1.3073	0.1227	CHOICE_VAGE	-1.1411	0.0323
CHOICE_VEHICLECLASS1	3.8437	1.1034	CHOICE_VEHICLECLASS1	5.4143	0.3937
CHOICE_VEHICLECLASS2	1.2740	1.4319	CHOICE_VEHICLECLASS2	3.4431	0.4058
CHOICE_CAPACITYCLASS1	-1.8475	1.0457	CHOICE_CAPACITYCLASS1	-0.5661	0.2697
CHOICE_CAPACITYCLASS2	-0.3714	0.8343	CHOICE_CAPACITYCLASS2	0.2698	0.2306
CHOICE_CAPACITYCLASS3	-0.7552	0.9288	CHOICE_CAPACITYCLASS3	0.5053	0.2314
CHOICE_BRANDCLASS1	0.3637	3.0305	CHOICE_BRANDCLASS1	0.6003	0.2338
CHOICE_BRANDCLASS2	-0.0063	3.0555	CHOICE_BRANDCLASS2	0.1862	0.2365
CHOICE_BRANDCLASS3	0.4519	3.5579	CHOICE_BRANDCLASS3	0.8607	0.2494
CHOICE_BRANDCLASS4	0.7596	3.0077	CHOICE_BRANDCLASS4	0.5420	0.2677
CHOICE_BRANDCLASS5	0.8170	3.0303	CHOICE_BRANDCLASS5	0.4219	0.3113
CHOICE_BRANDCLASS6	0.4681	2.9140	CHOICE_BRANDCLASS6	0.8437	0.2984
CHOICE_BRANDCLASS7	1.9331	2.8129	CHOICE_BRANDCLASS7	0.7659	0.3352
CHOICE_BRANDCLASS8	0.9907	3.0891	CHOICE_BRANDCLASS8	0.9379	0.2472
CHOICE_EXPERIENCE	-0.4684	0.2490	CHOICE_EXPERIENCE	-0.0004	0.0080
CHOICE_NCD0	-0.7702	0.5392	CHOICE_NCD0	-1.2750	0.1416
CHOICE_NCD10	-1.1398	0.5813	CHOICE_NCD10	-0.9052	0.1610
CHOICE_NCD20	-0.3001	0.6518	CHOICE_NCD20	-0.6248	0.1697
CHOICE_NCD30	-0.5398	1.0081	CHOICE_NCD30	-0.1376	0.2000
CHOICE_NCD40	1.1208	0.8804	CHOICE_NCD40	-0.3506	0.2167
RISK_INT	-1.5122	0.6850	RISK_INT	-1.5929	0.2689
RISK_AGEINSURED	-0.0036	0.0047	RISK_AGEINSURED	-0.0058	0.0021
RISK_SEXINSUREDF	0.0774	0.1040	RISK_SEXINSUREDF	-0.0384	0.0377
RISK_MARITALM	-0.1572	0.1253	RISK_MARITALM	0.0052	0.0423
RISK_VAGE	-0.0290	0.0081	RISK_VAGE	-0.0301	0.0028
RISK_VEHICLECLASS1	0.4113	0.5417	RISK_VEHICLECLASS1	0.7303	0.2463
RISK_VEHICLECLASS2	0.7770	0.5486	RISK_VEHICLECLASS2	0.6806	0.2484
RISK_CAPACITYCLASS1	-0.1764	0.2225	RISK_CAPACITYCLASS1	-0.0890	0.0747
RISK_CAPACITYCLASS2	-0.0573	0.1896	RISK_CAPACITYCLASS2	0.0117	0.0607
RISK_CAPACITYCLASS3	0.1319	0.1874	RISK_CAPACITYCLASS3	0.0238	0.0574
RISK_BRANDCLASS1	0.2291	0.3244	RISK_BRANDCLASS1	-0.0581	0.0804
RISK_BRANDCLASS2	0.2735	0.3306	RISK_BRANDCLASS2	0.0976	0.0811
RISK_BRANDCLASS3	0.3244	0.3276	RISK_BRANDCLASS3	-0.0493	0.0800
RISK_BRANDCLASS4	0.3766	0.3363	RISK_BRANDCLASS4	-0.0174	0.0851
RISK_BRANDCLASS5	-0.0736	0.3836	RISK_BRANDCLASS5	-0.0072	0.0972
RISK_BRANDCLASS6	0.0101	0.3928	RISK_BRANDCLASS6	-0.1369	0.1105
RISK_BRANDCLASS7	0.3875	0.3464	RISK_BRANDCLASS7	0.0413	0.0891
RISK_BRANDCLASS8	0.2203	0.3291	RISK_BRANDCLASS8	0.0264	0.0808
RISK_EXPERIENCE	0.0407	0.0576	RISK_EXPERIENCE	0.0004	0.0025
RISK_NCD0	0.0914	0.1402	RISK_NCD0	0.1735	0.0433
RISK_NCD10	-0.0224	0.1593	RISK_NCD10	0.2095	0.0469
RISK_NCD20	0.1414	0.1542	RISK_NCD20	0.0792	0.0508
RISK_NCD30	-0.0321	0.2016	RISK_NCD30	0.1491	0.0548
RISK_NCD40	0.2090	0.1990	RISK_NCD40	0.1551	0.0574
DEPENDENCE	2.6094	1.8370	DEPENDENCE	2.3427	0.3959
Log-likelihood	-677.5467		Log-likelihood	-6345.2810	
χ^2-statistic	1.9429		χ^2-statistic	41.4380	

Table 6: Estimation for the copula model by policyholder's age

	Young				Mid-age				Old			
	Choice - Logit		Risk - Probit		Choice - Logit		Risk - Probit		Choice - Logit		Risk - Probit	
	Estimate	StdErr	Estimate	StdErr	Estimate	StdErr	Estimate	StdErr	Estimate	StdErr	Estimate	StdErr
INT	14.9149	28.5980	-15.7072	11.3276	7.0591	0.5843	-1.7205	0.3116	5.8938	2.0565	-0.8307	0.8053
AGEINSURED	-0.1881	0.2885	0.0715	0.0770	0.0003	0.0077	-0.0063	0.0023	0.0079	0.0259	-0.0035	0.0091
SEXINSUREDF	0.4966	0.7645	-0.1292	0.2079	0.1374	0.1416	-0.0139	0.0374	0.2870	0.4526	-0.1326	0.1365
MARITALM	1.4848	0.9248	0.0918	0.2207	0.0193	0.1472	0.0019	0.0426	0.1099	0.8821	0.0784	0.3236
VAGE	-1.6363	0.2647	-0.0383	0.0163	-1.1673	0.0341	-0.0290	0.0029	-1.0592	0.0833	-0.0381	0.0085
VEHICLECLASS1	6.0740	24.6050	6.7029	15.7651	4.9163	0.3882	0.8900	0.2852	5.3790	0.9239	-0.1001	0.4259
VEHICLECLASS2	2.2485	24.6113	6.5637	15.7638	2.9218	0.4060	0.8925	0.2869	3.1614	0.9617	0.0247	0.4417
CAPACITYCLASS1	3.2396	9.7297	-0.2629	0.4644	-0.4966	0.2827	-0.0794	0.0756	-2.0214	0.6490	-0.7483	0.2558
CAPACITYCLASS2	4.4041	9.7141	-0.3668	0.4491	0.2905	0.2361	-0.0114	0.0613	-0.4924	0.5902	0.0140	0.1838
CAPACITYCLASS3	4.3235	9.7327	-0.4905	0.4536	0.5571	0.2381	0.0291	0.0582	-0.7475	0.5615	0.0106	0.1699
BRANDCLASS1	1.3910	2.0451	3.7747	17.3062	0.4597	0.2422	-0.0759	0.0821	1.1149	0.6528	0.0982	0.2474
BRANDCLASS2	0.5517	1.9850	3.8279	17.3061	0.0475	0.2460	0.1030	0.0828	0.9936	0.7053	0.0519	0.2538
BRANDCLASS3	-1.5361	2.0750	3.6523	17.3058	0.8349	0.2556	-0.0275	0.0813	0.9792	0.7291	-0.0207	0.2512
BRANDCLASS4	0.7848	2.1130	3.8128	17.3044	0.6069	0.2777	-0.0215	0.0867	0.0434	0.7473	0.2518	0.2707
BRANDCLASS5	0.2145	2.6746	3.3926	17.3009	0.4412	0.3216	-0.0048	0.0979	0.5142	0.9860	-0.2577	0.3773
BRANDCLASS6	1.5512	2.3774	3.2413	17.3017	0.8559	0.3127	-0.1369	0.1124	0.2068	0.7724	0.0180	0.3503
BRANDCLASS7	0.8215	2.0266	3.5119	17.3024	0.9922	0.3655	0.0874	0.0899	0.7388	0.9733	-0.3388	0.3247
BRANDCLASS8	1.2780	2.1338	3.6964	17.3044	0.9387	0.2577	0.0321	0.0823	1.0122	0.6680	-0.0153	0.2541
EXPERIENCE	-0.0401	0.1632	-0.0743	0.0490	-0.0067	0.0075	-0.0017	0.0023	-0.0204	0.0108	0.0007	0.0036
NCD0	-5.8009	24.7511	3.6686	17.2162	-1.2035	0.1438	0.1547	0.0437	-1.0172	0.4347	0.0416	0.1637
NCD10	-4.9758	24.7790	3.3156	17.2133	-0.8806	0.1665	0.1957	0.0475	-1.0801	0.4682	0.1894	0.1639
NCD20	-6.4980	24.8019	3.4956	17.2123	-0.5883	0.1760	0.0903	0.0510	-0.3124	0.4623	-0.0366	0.1687
NCD30	5.0232	172.0251	4.4442	17.2032	-0.0769	0.1991	0.1223	0.0558	-1.2143	0.5832	0.2454	0.1680
NCD40	-7.4539	24.8423	4.5520	17.2088	-0.2180	0.2239	0.1491	0.0582	0.0387	0.6330	0.1716	0.1817
DEPENDENCE	2.4440	1.5863			2.3547	0.4192			2.3425	1.2913		
Log-likelihood	-194.6840				-6125.2930				-675.4912			
χ^2-statistic	2.0110				37.4140				3.0176			

6 Summary and Concluding Remarks

In this paper, we studied adverse selection in insurance explicitly assuming that an individual's risk and risk aversion are the two characteristics that affect insurance purchase decision. Building on previous works, we developed a theoretical model of insurance screening in a two-dimensional information framework. We believe that the degree of asymmetric information and consequently adverse selection vary across different subgroups of policyholders. Thus, we examined two cases of the pattern of asymmetric information: either the policyholder is privately informed of his risk and risk aversion, or he only knows his attitude toward risk. In the former case, insurance buyers of different risk types can be successfully separated in the equilibrium as in the one-dimensional model, if the effect of risk aversion is not substantial. Otherwise, policyholders of different risk levels are partially pooled. In the latter consideration, we discussed the implications of the potential correlation between risk and risk aversion. When they are distributed independently, knowing risk aversion amounts to no information advantage of one's riskiness, therefore the equilibrium pools all individuals in a single contract. When risk and risk aversion are correlated and buyers can infer their riskiness from risk aversion, again adverse selection arises and the equilibrium is a separating one.

In the empirical investigation, we examined coverage-risk relationship in the Singapore automobile insurance market using a cross-sectional observation from a major insurer. We used the Frank copula to jointly model the type of policy coverage and accident occurrence, with the dependence parameter of the copula to help describe any evidence of private information. The copula model allows for the flexibility in examining each component in the coverage-risk setup and captures both linear and non-linear relationships. When analyzing the entire portfolio of policyholders from the insurer, we found significant positive association between policyholder's choice of coverage and occurrence of accidents, indicating the presence of adverse selection. Furthermore, we disaggregated the population by policyholder's characteristics and investigated the equilibria predicted by the theoretical insurance screening model. Firstly, we examined the coverage-risk relation according to a policyholder's driving experience. There was no significant relationship found for the group of inexperienced drivers but a significant positive association found for experienced drivers. This can be explained by beginning drivers' lack of knowledge of their riskiness and the relation between risk and risk aversion. Secondly, we looked at the coverage-risk relationship by policyholder's age. We found non-linear age effects on the relationship: a positive association exists for mid-aged drivers and the relation becomes statistically insignificant for young and old policyholders. This appears to be consistent with the theoretical predictions, if one takes age as an important determinant of risk aversion and accepts that as one gets older, the effect of risk aversion in one's insurance purchase decision becomes stronger.

We note that our analysis has focused on the coverage-risk relationship when examining the behaviors of policyholders with private information. Based on a cross-sectional observation, the analysis is limited to a combined result of adverse selection and moral hazard. However, as suggested by Abbring et al. (2003) and Abbring et al. (2003), dynamic data could help distinguish adverse selection from moral hazard. We leave this as one of the future research directions to pursue. The emphasis of this work is on the adverse selection in a two-dimensional information context, where both the level of risk and the degree of risk aversion are different across policyholders. Our contribution to the current literature is the two-dimensional information insurance screening setup and the copula regression model in the empirical analysis. The evidence of adverse selection found in the Singapore market supplements the existing findings on the empirical testing of asymmetric information in auto insurance, and more importantly, the economic model built in our work provides theoretical foundations for such empirical findings.

References

Abbring, J., P. A. Chiappori, and J. Pinquet (2003). Moral hazard and dynamic insurance data. *Journal of the European Economic Association 1*(4), 767–820.

Abbring, J., J. Heckman, P. A. Chiappori, and J. Pinquet (2003). Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish? *Journal of the European Economic Association 1*(2-3), 512–521.

Boyer, M. and G. Dionne (1989). An empirical analysis of moral hazard and experience rating. *Review of Economics and Statistics 71*(1), 128–134.

Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *Econometrics Journal 7*(2), 566–584.

Chiappori, P. A. (2000). Econometric models of insurance under asymmetric information. In G. Dionne (Ed.), *Handbook of Insurance*, pp. 365–394. Kluwer Academic Publishers.

Chiappori, P. A., B. Jullien, B. Salanié, and F. Salanié (2006). Asymmetric information in insurance: General testable implications. *RAND Journal of Economics 37*(4), 783–798.

Chiappori, P. A. and B. Salanié (1997). Empirical contract theory: The case of insurance data. *European Economic Review 41*(3-5), 943–950.

Chiappori, P. A. and B. Salanié (2000). Testing for asymmetric information in insurance market. *Journal of Political Economy 108*(1), 56–78.

Chiappori, P. A. and B. Salanié (2003). Testing contract theory: A survery of some recent work. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications. Eight World Congress*, Volume 1, pp. 115–149. Cambridge University Press.

Cohen, A. (2005). Asymmetric information and learning: evidence from the automobile insurance market. *Review of Economics and Statistics 87*(2), 197–207.

Cohen, A. (2008). Asymmetric learning in repeated contracting: An empirical study. *NBER Working paper*.

Cohen, A. and P. Siegelman (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance 77*(1), 39–84.

de Donder, P. and J. Hindriks (2009). Adverse selection, moral hazard and propitious selection. *Journal of Risk and Uncertainty 38*(1), 73–86.

de Meza, D. and D. Webb (2001). Advantageous selection in insurance markets. *RAND Journal of Economics 32*(2), 249–262.

Dionne, G. and N. Doherty (1992). Adverse selection in insurance markets: A selective survey. In G. Dionne (Ed.), *Contributions to Insurance Economics*, pp. 97–140. Kluwer Academic Publishers.

Dionne, G., N. Doherty, and N. Fombaron (2000). Adverse selection in insurance markets. In G. Dionne (Ed.), *Handbook of Insurance*, pp. 185–244. Kluwer Academic Publishers.

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.

Dionne, G., M. Maurice, J. Pinquet, and C. Vanasse (2005). The role of memory in long-term contracting with moral hazard: Empirical evidence in automobile insurance. *Working Paper*.

Dionne, G., P.-C. Michaud, and M. Dahchour (2010). Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from France. *Working Paper*.

Dionne, G. and C. Vanasse (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics 7*(2), 149–165.

Frees, E., P. Shi, and E. Valdez (2009). Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin 39*(1), 165–197.

Frees, E. and E. Valdez (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association 103*(484), 1457–1469.

Frees, E. and P. Wang (2005). Credibility using copulas. *North American Actuarial Journal 9*(2), 31–48.

Frees, E. and P. Wang (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics 38*(2), 360–373.

Hemenway, D. (1990). Propitious selection. *Quarterly Journal of Economics 105*(4), 1063–1069.

Hemenway, D. (1992). Propitious selection in insurance. *Journal of Risk and Uncertainty 5*(3), 247–251.

Israel, M. (2006). Where Is all the hidden information hiding? Evidence from automobile insurance panel data.

Israel, M. (2007). Do we drive more safely when accidents are more expensive? Identifying moral hazard from experience rating schemes. *Working Paper*.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall.

Jullien, B., B. Salanié, and F. Salanié (2007). Screening risk-averse agents under moral hazard: Single-crossing and the CARA case. *Economic Theory 30*(1), 151–169.

Kim, H., D. Kim, S. Im, and J. Hardin (2009). Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. *Journal of Risk and Insurance 76*(2), 343–366.

Koufopoulos, K. (2007). On the positive correlation property in competitive insurance markets. *Journal of Mathematical Economics 43*(5), 597–605.

Landsberger, M. and I. Meilijson (1994). Monopoly insurance under adverse selection when agents differ in risk aversion. *Journal of Economic Theory 63*(2), 392–407.

Li, C., C. Liu, and J. Yeh (2007). The incentive effects of increasing per-claim deductible contracts in automobile insurance. *Journal of Risk and Insurance 74*(2), 441–459.

Morin, R. and A. Suarez (1983). Risk aversion revisited. *Journal of Finance 38*(4), 1201–1216.

Nelsen, R. (2006). *An Introduction to Copulas.* Springer.

Pålsson, A. (1996). Does the degree of relative risk aversion vary with household characteristics? *Journal of Economic Psychology 17*(6), 771–787.

Prieger, J. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics 17*(4), 367–392.

Puelz, R. and A. Snow (1994). Evidence on adverse selection: equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy 102*(2), 236–257.

Purcaru, O. and M. Denuit (2005). Dependence in dynamic claim frequency credibility models. *ASTIN Bulletin 33*(1), 23–40.

Richaudeau, D. (1999). Automobile insurance contracts and risk of accident: an empirical test using French individual data. *The Geneva Papers on Risk and Insurance - Theory 24*(1), 97–114.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *The Quarterly Journal of Economics 90*(4), 629–649.

Saito, K. (2006). Testing for asymmetric information in the automible insurance market under rate regulation. *Journal of Risk and Insurance 73*(2), 335–356.

Shi, P. and E. Frees (2010). Long-tail longitudinal modeling of insurance company expenses. *Insurance: Mathematics and Economics 47*(3), 303–314.

Smart, M. (2000). Competitive insurance markets with two unobservables. *International Economic Review 41*(1), 153–169.

Smith, M. (2003). Modeling Selectivity Using Archimedean Copulas. *Econometrics Journal 6*(1), 99–123.

Stiglitz, J. (1977). Monopoly, non-linear pricing, and imperfect information: the insurance market. *Review of Economic Studies 44*(3), 407–430.

Wambach, A. (2000). Introducing heterogeneity in the Rothschild-Stiglitz model. *Journal of Risk and Insurance 67*(4), 579–591.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory 16*(2), 167–207.

Zimmer, D. M. and P. K. Trivedi (2006). Using trivariate copulas to model sample selection and treatment effects. *Journal of Business and Economic Statistics 24*(1), 63–76.