Can Purchasing Records Predict Risk?

Larry Y. Tzeng

Professor, Department of Finance,

National Taiwan University, Taipei, Taiwan

Kili C. Wang*

Associate Professor, Department of Insurance,

Tamkang University, Taipei, Taiwan

Abstract

In the paper, we propose a new method to examine adverse selection against advantageous selection, and our empirical evidence supports the existence of adverse selection as shown by Rothschild and Stiglitz (1976). By using a unique tracking sample with 216,942 observations based on Taiwan's comprehensive automobile insurance market, we find that the past purchasing records of insurance coverage can predict the occurrence of risk in the future. We also find that controlling past records of the insured is only able to mitigate but not eliminate the asymmetric information problems in the market. We further provide evidence to show that both moral hazard and adverse selection could co-exist in the market. Since past records could control adverse selection but not moral hazard, past records may have a limited effect in terms of controlling asymmetric information with both adverse selection and moral hazard.

* Corresponding Author. Department of Insurance, Tamkang University, Taipei, 104, Taiwan. E-mail address: kili@mail.tku.edu.tw, tel: 886-2-2621-5656 ext: 3333.

Can Purchasing Records Predict Risk?

Introduction

The adverse selection theory has been one of the cornerstones in insurance both in academia and in practice. Rothschild and Stiglitz (1976) have shown that the insurance market under adverse selection could settle on a separating equilibrium where high risks purchase high coverage at a high premium rate and low risks purchase low coverage at a low premium rate. Hence, the risk and the coverage are positively correlated.

Following the theoretical studies, many empirical papers have also found evidence of adverse selection by examining whether there exists a positive relationship between risk and coverage[1], for example, Cardon and Hendel (2001) in the health insurance market of the US, Finkelstein and Poterba (2002) in the annuity market in the UK, and Cohen (2005) in the automobile insurance market in Israel. However, many empirical papers have failed to find empirical evidence for the positive relationship between risk and coverage to support the existence of adverse selection in the real insurance business[2]. Some of the studies have even found a

---

[1] It is worth pointing out that both moral hazard and adverse selection predict a positive relationship between risk and coverage as discussed by Chiappori and Salanie (2000).

[2] Some papers have found that the choice of the coverage and the occurrence of the risk are not significantly correlated, such as Chiappori and Salanie (1997), Richaudeau (1999), Chiappori and Salanie (2000) and Dionne, Gourieroux and Vanasse (2003) in the automobile insurance market in

significantly negative relationship between risk and coverage, such as Cawley and Philipson (1999) in the life insurance market in the US, Finkelstein and McGarry (2006) in the long-term care insurance market in the US, Davidoff and Welke (2004) in the reverse mortgage insurance market in the US, Gronqvist (2004) in the dental care insurance market in Sweden, and Fang, Keane and Silverman (2005) in the medigap insurance market in the US.

To explain why the choice of the coverage and the occurrence of the risk are negatively correlated, de Meza and Webb (2001) have provided an intriguing theory which they refer to as advantageous selection. de Meza and Webb (2001) assume that the insurer cannot observe the risk preference of the insured and the effort of the insured in terms of loss prevention. They further show that highly risk-averse individuals expend more effort on reducing the loss probability to become low-risk insured, whereas low risk-averse individuals expend less effort in order to reduce the loss probability and become high-risk insured. Since the risk preference of the individual is private information, the insurer screens the insured by offering high coverage to attract highly risk-averse individuals while offering low coverage to attract low risk-averse individuals. Therefore, asymmetric information in terms of the risk preference could make the market equilibrium settle on high risk with low

---

France; Dionne, Gourieroux and Vanasse (2001) in the automobile insurance market in Canada, and Saito (2006) in the automobile insurance market in Japan.

coverage and on low risk with high coverage.

It is very important to recognize that both Rothschild and Stiglitz (1976) and de Meza and Webb (2001) adopt one-period models. Thus, if the insurance market settles on a separating equilibrium, the insurer should in the next period have full information on the risk type or risk preference of the insured. When the equilibrium of an insurance market under adverse selection is a separating equilibrium as shown by Rothschild and Stiglitz (1976), the insurance companies can predict the risk types of those who continue to be insured in the next period since the market equilibrium has screened the policyholders. The insured who purchases a high coverage in the last period will be a high risk type individual. On the other hand, when the equilibrium of an insurance market under advantageous selection is a separating equilibrium as shown by de Meza and Webb (2001), the insurance companies could also predict the preference of those continuing to be insured in the next period since the separating equilibrium has screened the degree of risk preference. In contrast to Rothschild and Stiglitz (1976), de Meza and Webb (2001) predict that the insured who purchases a high coverage contract in the last period will be a low-risk individual.

In this paper, we intend to examine whether purchasing records do in fact contain valuable information to predict the risk of the insured. The method we propose could be treated as a new methodology to distinguish adverse selection from advantageous selection. In the literature, most papers identify adverse selection or advantageous selection based on the relationship between the risk and coverage in the same year. In this paper, we employ the relationship between purchasing records (coverage) in the previous year and claim records (risk) in the current year to examine whether adverse selection rather than advantageous selection exists in the market.

Using purchasing records to test the existence of adverse selection could further separate adverse selection from moral hazard. In the literature, many papers test the correlation between risk and coverage in the same year to identify the existence of adverse selection. However, as noted by Chiappori and Salanie (2000), both moral hazard and adverse selection predict a positive correlation between the choice of the coverage and the occurrence of the risk. Thus, the method we propose in this paper could also serve to provide evidences for the existence of adverse selection which cannot be explained by moral hazard. We investigate the relationship between the occurrence of the risk in this year and the purchasing records in the previous year. Moral hazard predicts that higher coverage, less effort on loss prevention, and higher risk. Thus, the occurrence of the risk this year should be correlated to the coverage in this year rather than the previous year purchasing records. Of course, it should be pointed out that finding evidence to support the existence of adverse selection does not automatically reject the possible existence of moral hazard.

While the separating equilibrium is a screening mechanism, it needs to be asked whether it offers information regarding the risk type of the insured or information regarding the risk preference of the insured. This can be judged by observing the relationship between the past purchasing records and the probability of risk occurring in this period. If there is a significant positive relationship between past purchasing records and the probability of claims occurring in this period, the screening mechanism could offer a chance to predict the risk of the insured as deduced by Rothschild and Stiglitz (1976). On the other hand, if there is a significant negative relationship between past purchasing records and the probability of claims occurring in this period, the screening mechanism could offer a chance to predict the risk preference of the insured and supports de Meza and Webb (2001).

The other question we ask in this paper is whether the asymmetric information could be partially reduced if the insurance company were able to observe the past purchasing records of the insured. We hypothesize that past purchasing records should be able to reduce the asymmetric information caused by either advantageous selection as noted by de Meza and Webb (2001) or adverse selection as described by Rothschild and Stiglitz (1976). Because policyholders have been screened under a separating equilibrium, past purchasing records contain information related to the risk types or risk preferences of individuals and should mitigate the problem of asymmetric information. However, we hypothesize that past purchasing records may not be able to reduce the asymmetric information caused by moral hazard, since moral hazard is induced by the coverage of the current insurance policy rather than the coverage of past insurance policies. Hence, after the past purchasing records are controlled for, the asymmetric information problem may still exist, and only its severity will have been reduced if moral hazard also exists in the market.

Our target is to investigate the asymmetric information problems in the comprehensive automobile insurance[3] market for Taiwan. We believe that the market provides a natural experiment for us to examine whether the purchasing records could predict risk and reduce asymmetric information. Before 1999, the Taiwan comprehensive automobile insurance market consisted of two types of insurance coverage: type A and type B. Both types A and B can be treated as "all-risk" coverage, since type A is exactly an all-risks basis insurance coverage and type B covers all risks except unknown risks. Due to the deteriorating loss ratio, in 1999 insurance companies launched type C coverage which covers the loss only if the automobile accident involves more than one car[4]. We believe that the launching of type C

---

[3] In Taiwan, there are three different types of comprehensive coverage: A, B, and C. Type A covers all kinds of collision and non-collision losses, including those caused by missiles or falling objects, fire, explosion, windstorm, intentional body damage, malicious mischief, and any unidentified reasons other than the exclusions in the policy. Type B covers all the areas for type A but excludes the non-collision losses caused by intentional body damage, malicious mischief, and any unidentified reasons. Type C covers only damage in a collision involving two or more vehicles. Collision losses caused by hitting other objects—such as a telephone pole, a tree, or a building—and non-collision losses that used to be covered under types A and B are specifically excluded from type C.

[4] In this paper, from the viewpoint of the risk the insurance contract covers, we separate those

coverage provides us with an opportunity to examine how insurance companies screen policyholders by offering product menu in the market.

Several papers in the literature have already provided evidence to support the existence of asymmetric information in Taiwan's comprehensive automobile insurance market. Wang (2006a, 2006b) found that, in a sample of type A and type C coverage as well as a sample of type B and type C coverage, the relationship between the coverage and the risk is positive and might support the existence of adverse selection and/or moral hazard. Furthermore, Li et al. (2007)[5] and Wang et al. (2008)[6] also found evidence to support the existence of adverse selection and/or moral hazard in the market. The existence of adverse selection and/or moral hazard could provide us with a unique opportunity to examine whether past purchasing records are able to control adverse selection and reduce asymmetric information.

Hence, our first purpose is to determine whether the past purchasing records could reveal some valuable information regarding the risk of the insured. We then try to identify whether an asymmetric information problem exists in the market and whether it can be eliminated or mitigated by controlling the past records. Furthermore, if the asymmetric information is only mitigated by past purchasing records, we try to

---

contracts according to whether the contract covers all risks or not. Hence, type A and B contracts are treated as high coverage contracts because they cover all risks, and a type C contract is treated as a low coverage contract because it only covers the risk of accidents involving more than one car.

5 They define coverage by the deductible level within the type B contract: the contracts with a low deductible are high coverage contracts, and the contracts with a high deductible are low coverage contracts.

6 Our definition of coverage is consistent with their paper.

provide some evidence to explain the phenomena.

Our data provide us with a unique advantage in examining the relationship between past records and asymmetric information. In the literature, relatively few papers[7] have access to information on past purchasing records. Since our data comprise a tracking sample, we are not only able to have past claim records, but we are also able to have access to the past purchasing records from the early stages of the launching of the type C contract. We started collecting the data immediately after the type C coverage was launched from the year 2001 to the year 2006. We traced the past purchasing records of the previous year[8] for the contracts which continued for at least one year. Hence, the data based on those continued contracts extends from the year 2002 to the year 2006, and we have used the pooled data with 216,942 observations to perform our tests.

By pooling these data, we find that the previous year's purchasing records of the insured could reveal valuable information regarding risk type. The insured have a higher probability of filing a claim this year if they chose a high coverage contract last year. The empirical evidence strongly supports the view that past purchasing records can increase the prediction power of loss probability. This is consistent with

---

[7] Dionne et al. (2006) use data from past purchasing records to try to disentangle moral hazard from the other asymmetric information phenomena.

[8] We have also already adopted the test which uses past purchasing records for more than one year. We find that all the empirical findings using more than one year of records are similar to those using one year of past purchasing records. Hence, in this paper, we only trace records for one year.

Rothschild and Stiglitz (1976) and also supports adverse selection rather than advantageous selection.

Not surprisingly, we also find that asymmetric information does exist in the early stages of this market. For type A and type B contracts versus the type C contract, the relationship between the choice of coverage and the occurrence of the risk is found to be significantly positive and supports the existence of adverse selection and/or moral hazard. This finding is consistent with Wang et al. (2008), Wang (2006a, 2006b), and Li et al. (2007), who also found that asymmetric information exists in this early stage of Taiwan's automobile insurance market.

Furthermore, we find that, after controlling past records (both the claim records and the purchasing records), the conditional correlation between the choice of coverage and the occurrence of the risk is still significantly positive, but the magnitude of the correlation becomes smaller. Our empirical evidence generally supports the view that the past purchasing records cannot fully eliminate the asymmetric problems if moral hazard and adverse selection co-exist in the market. If the market is haunted by moral hazard, past records could eliminate adverse selection rather than moral hazard. Controlling past records may have a limited effect on mitigating the asymmetric information in the market.

To support the above assertion, we further investigate whether moral hazard exists in the market. The methodology we employ to test for the existence of moral hazard is inspired by Abbring et al. (2003). They argued that the insured who has already filed a claim in the previous year will be charged a high premium this year and he/she will change his/her purchasing decision from high coverage to low coverage. If the behavior of the insured is unobservable information for the insurer, i.e., there is a moral hazard problem, after the higher premium is charged, he/she will change to a lower coverage contract and pay more attention. Therefore, his/her probability of making a claim this year will be lower and negatively correlated with the occurrence of the claim in the previous year. However, Abbring et al. (2003) did not find such evidence.

Abbring et al. (2003) only test the relationship between the previous year's claims and this year's claims. In this paper, we will not only examine both the relationship between the insured's previous year's claims and this year's claims and the relationship between the insured's previous year's claims and this year's choice of coverage, but we will also further examine the trend in terms of the choice of coverage and the claims in this year for the insured who had purchased high coverage and had claimed in the previous year. We find that the insured's previous year's claims and this year's claims are positively correlated while the insured's previous year's claims and this year's choice of coverage are negatively correlated.

Furthermore, we find that the insured who purchased high coverage and filed a claim in the previous year is even less likely to file a claim this year than the insured who purchased low coverage and filed a claim in the previous year. Moreover, we find that the insured who purchased high coverage and filed a claim in the previous year is even less likely to purchase a high-coverage contract than the insured who purchased low coverage and filed a claim in the previous year. This can more clearly describe the behavior of moral hazard. Thus, our empirical evidence supports the existence of moral hazard and indirectly explains why past purchasing records cannot fully eliminate asymmetric information in the market.

Our empirical findings contribute to the literature in three ways. First, we propose a new method to examine the existence of adverse selection and find that past purchasing records contain valuable private information regarding the risk type of the insured. This finding supports the prediction of Rothschild and Stiglitz (1976) and cannot be explained by moral hazard. Secondly, we find that, after controlling for the past purchasing records, there still exists a positive conditional correlation between coverage and risk. We assert that past records could control adverse selection but not moral hazard. Third, we further test the dynamic behavior of the insured and, in addition to adverse selection, indirectly find evidence to support the view that moral hazard also exists, and thereby further strengthen our argument.

This paper is organized as follows. The first section consists of the Introduction. In the second section we describe the data, and the methodology we have adopted is described in the third section. The empirical results are then presented in the fourth section, and the final section provides the Conclusion.

Data

Our data are obtained from one large insurance company in Taiwan. Since the loss rate and loss ratio of that company are close to that of the market as a whole, we believe that the data from this insurance company are representative of the industry. The dependent variables comprise the choice of the coverage and the occurrence of the claim. The variables representing the information used by the insurer in underwriting and/or pricing include the content and the distribution channel of the contract, and the characteristics of the insured. We also collect both the claim records and the purchasing records. The definitions of all variables are given in Table 1.

The sample period extends from the year 2001 to the year 2006. In each sample year we estimated, we traced one year of past records. Hence, the contracts that this paper investigates are those of the insured who continued in this insurance company for at least one year. Accordingly, our sample contains five sample periods which are the years from 2002 to 2006. We pool them together to implement our empirical analysis and control them by means of four year dummy variables: *year_2002, year_2003, year_2004*, and *year_2005*, with the sample for the year 2006 serving as the reference group. The percentage for each sample period can be observed from *year_2002, year_2003, year _2004,* and *year_2005* in Table 2, with the growth trend of the ongoing contracts in this market that are owned by this insurance company being displayed there. A total of 43.68% of the continued contracts are high coverage type A and type B contracts.

Most of the insured are good drivers from the point of view of experience rating. The percentage of insured whose bonus-malus coefficients are under 0.9 is over 64%, while the percentage of insured whose bonus-malus coefficients are above 0.9, but under 1, is 31.81%. Some 24% of those who continue to be insured have filed a claim in the current year, and the average number of claims is 0.29. Hence, the average number of claims for those who are insured who have already filed a claim this year is 1.29.

By observing the previous year's purchasing records and the structure of different types of contracts this year, we find that there is no dramatic structural change in the percentages of contract coverage year by year, although there is a gradually decreasing trend in high coverage contracts.

With previous year purchasing and claim records, we create a total of three variables to filter information concerning the characteristics of the insured. These variables include one variable for past purchasing records, one variable for claims in the previous year, and one multiple term created by the purchasing and claim records in the previous year[10].

---

[9] The maximum number of claims to occur in our sample is eight in one year.

[10] The term *clm_b1*coverage_b1* refers to the insured who purchased high coverage and who had also made a claim in the previous year

Methodology

In our empirical work, we first investigate whether past purchasing records contain valuable information concerning risk. Hence, we test the significance of purchasing records in explaining accidents.

Through observing our data set, we find that the numbers of claims for the insured who have previously filed claims vary quite significantly. We believe that the degree of risk could be heterogeneous among those who filed different numbers of claims. Hence, to treat each insured who has filed a claim by ignoring his or her claim number could lead to bias while we estimate the probability of a claim as representing the degree of the insured's risk.

To avoid the above concern, we consider that the risk degree of risk could vary with the number of accidents. Hence, we use the negative binomial model instead of a Probit regression model to estimate the probability of risk as was the case in Lemaire (1989) and Dionne and Vanasse (1989).

In the negative binomial model, the regression component is introduced in the negative binomial model:

$$prob(Y_i = y \mid X_{1i}) = \frac{\Gamma(y + 1/\alpha)}{y!\Gamma(1/\alpha)} \frac{\left[\alpha \exp(X_{1i}\beta)\right]^y}{\left[1 + \alpha \exp(X_{1i}\beta)\right]^{y+1/\alpha}} \qquad (1)$$

The above expression represents a negative binomial distribution with

$E(Y_i \mid X_i) = \exp(X_{1i}\beta)$ and $Var(Y_i \mid X_{1i}) = \exp(X_{1i}\beta)(1 + \alpha \exp(X_{1i}\beta))$, and

$Y_i = clmno_i$. The vector X_{1i} contains the explanatory variables[11] which could affect the number of accidents involving the insured. $Y_i = clmno_i$ is the real number of accidents that occurred. α and β are the estimated coefficients for the exponential term in the distribution function and each corresponding characteristic variable.

Note that we treat a policy as one with higher coverage if the policy provides a broader coverage of accident items in this paper. We treat both type A and type B contracts as high coverage because both of them involve all-risks type coverage, whereas type C contracts cover collision only. However, it should be recognized that both type A and type B contracts include deductibles, while type C contracts do not. Thus, we further adjust the deductible while we define the claim to calculate $clmno_i$[12].

Meanwhile, the type C contract only covers the collision accident, while the type A and B contracts also cover the accident beyond the collision. To avoid the concern over spurious correlation as mentioned by Chiappori and Salanie (2000), for all types of contracts, we define the accident as only a claim involving a collision so as to have a consistent base among different contracts.

---

[11] The vector X_1 contains the past records, the multiple term for the past purchasing and claim records (*clm_b1*coverage_b1*), the demographic variables of the insured, and the control variables in Table 1.

[12] We regard the deductible for all of the type C contracts as being NT$3,000. The way we adjust the deductibles between high coverage contracts (i.e., type A and B contracts) and low coverage contracts (i.e., type C contracts) involves defining an accident by a claim for a type C contract as being NT$3,000 higher than a claim for a type A or type B contract.

The parameters α and β in equation (1) are estimated by the maximum likelihood method. When we seek to determine whether there is valuable information regarding the risk revealed by the past purchasing records or not, the key coefficient we test is the estimated coefficient of the past purchasing record variables, i.e., *coverage_b1*.

If the coefficient of the previous year purchasing record variable (*coverage_b1*) is significantly positive, this means that the higher coverage that the insured chose in the previous year may result in the insured having more accidents in the current year. This evidence proves that purchasing records contain valuable information regarding the risk type of the insured and supports what Rothschild and Stiglitz (1976) predicted. If it is significantly negative, it means that the higher coverage that the insured chose in the previous year may result in the insured having fewer accidents in the current year. This evidence proves that the purchasing records contain valuable information regarding the risk preference of the insured and supports what de Meza and Webb (2001) predicted.

The second intention of this paper is to test the existence of the asymmetric information in this market, and to further test whether this asymmetric information could be mitigated by controlling the past records. Hence, we apply two models to test the asymmetric information We do not control the past record variables[13] in Model 1, whereas we control the past record variables in Model 2. The asymmetric information phenomena are tested based on the conditional correlation between risk and coverage. Similar to the methodology used by Dionne, Gourieroux and Vanasse (2001), the conditional correlation is obtained by using the two-stage method.

In the first stage, we use a negative binomial model associated with all the exogenous variables concerned with the characteristics of the insured to calculate the expected number of accidents ($cl\hat{m}no$) of the insured. The second stage is then a Probit on the coverage choice of the insured:

$$\begin{aligned} &\mathrm{Prob}(coverage_i = 1 \mid cl\hat{m}no_i, clmno_i, X_{2i}) \\ &= \Phi(\beta_1 cl\hat{m}no_i + \beta_2 clmno_i + X_{2i}\beta_3) \end{aligned} \qquad (2)$$

where $coverage$ is the dummy variable of coverage choice which is defined in Table 1. The regressors include the estimated number of accidents from the first stage ($cl\hat{m}no$), the number of accidents this year ($clmno_i$), as well as the vector X_{2i} which contains the factors that affect the coverage choice of the insured[14]. β_1, β_2, and β_3 are the estimated coefficients of $cl\hat{m}no$, $clmno_i$ and X_{2i}, respectively.

In this Probit regression, among all the estimated coefficients, β_2 is the key coefficient we use to test the asymmetric information. If β_2 is significantly positive,

---

[13] This includes the past record variables in Table 1 and the multiple term *clm_b1*coverage_b1*.

[14] In Model 2, the variables included in X_{2i} are all the explanatory variables contained in X_{1i} as well as the variable of *lnprem*. In Model 1, the variables in X_{2i} not includes all the past record variables.

it is evidence that adverse selection and/or moral hazard exists in the market. If β_2 is significantly negative, it is evidence that advantageous selection exists in the market. Hence, in Model 1, we observe the significant level and the sign of β_2 to infer whether and what asymmetric information exists in the market. We further intend to test whether the asymmetric information could be mitigated through controlling the past records. In Model 2 which controls the previous year's past records in the first stage, if β_2 is significantly different from zero in Model 1 and becomes insignificantly different from zero (significantly different from zero but the value turns smaller), we can infer that the asymmetric information of this market can be eliminated (be mitigated) by controlling the past records of the insured.

If the asymmetric information in this market is only mitigated instead of being eliminated after the previous year's past records are controlled for, the third intention of this paper is to explain the phenomena by providing evidence to support the view that moral hazard may also exist in this market.

We follow the ideas of Abbring et al. (2003). we have examined the relationship between the past claims and this year's claims using the negative binomial regression model (1). To more directly test the existence of moral hazard, we further investigate the relationship between the past records and this year's purchasing decisions. Hence, we build the following Probit regression for the purchasing decisions:

$$\text{Prob}(\text{cov}\,erage_i = 1 \mid X_{3i}) = \Phi(X_{3i}\gamma) \qquad (3)$$

where the definition of $\text{cov}\,erage$ is the same as that in regression (2). The vector X_{3i} contains the past records of *clm_b1*, *coverage_b1* and their multiple term *clm_b1*coverage_b1*, as well as the factors affecting the coverage choice of the insured[15]. γ is their correspondent coefficient vector. In this regression, the key variables that need to be observed are those coefficients of *clm_b1*, *coverage_b1*, and *clm_b1*coverage_b1*.

We predict that the coefficient of *clm_b1* could be negative because of the punishment of experience rate, and the coefficient of *coverage_b1* should be positive if the past purchasing records reveal the information regarding the insured's risk type. Meanwhile, if the moral hazard exists, the insured who has purchased high coverage and has already filed a claim in the previous year will change to a lower coverage contract in the current year due to being punished by the bonus-malus coefficient. Meanwhile, he/she will reduce his/her claims in the current year. Hence, we predict that the coefficient of *clm_b1*coverage_b1* in the negative binomial regression (1) should be significantly negative, and that the coefficient of *clm_b1*coverage_b1* in the Probit regression (3) should be significantly negative, too.

---

[15] In words, all the variables are already included in X_{2i}.

Empirical results

Whether or not valuable information is revealed by the past purchasing records can be observed from the negative binomial model which is regressed on the number of claims. The empirical results are presented in Table 3. The estimated coefficient of *coverage_b1* is significantly positive. From this outcome, we can infer that the previous year's purchasing records do reveal important information regarding the risk of the insured. The higher the coverage that the insured chose in the previous year could result in the insured being involved in more accidents this year. This evidence does support the prediction of Rothschild and Stiglitz (1976). By observing the choice of coverage of the insured in the previous year, the insured who chose a contract with high coverage could signal that he or she is a high risk type, hence, and could be involved in more accidents next year.

Furthermore, there are also some by-products that can be observed from the outcomes in Table 3. First, the coefficient of the past claim records is significantly positive. These outcomes reveal that the risk type of the insured cannot be changed over time, and thus the one who is currently a high risk type may always be a high risk type. This makes the correlation between the previous year's claims and this year's claims positive. Hence, this outcome could be regarded as evidence of adverse selection. Secondly, the coefficient of *clm_b1*coverage_b1* is significantly negative.

This means that the insured who purchased high coverage in the previous year but had filed a claim would be less likely to file a claim this year. Why is that? Because of the experience rating system, the insured who filed a claim in the previous year could be charged a higher premium and therefore may change his or her choice of coverage and adopt low coverage instead. The low insurance coverage could further provide a stronger incentive for the insured to expend more efforts on preventing the occurrence of the risk, which implies the existence of moral hazard.

The empirical results for the conditional correlation between risk and coverage based on the two-stage method are listed in Table 4. In Model 1, before we control any past records, the coefficient β_2 is significantly positive. This is evidence that adverse selection and/or moral hazard exist in this market. Such a finding is consistent with those presented in Table 3.

In Model 2, after controlling the records of the past year, the coefficient β_2 is still significantly positive. However, compared with its value in Model 1, the value of coefficient β_2 declines from 0.7606 to 0.4283. This outcome implies that the asymmetric information phenomenon still exists after past records are controlled. In spite of this, with the help of the information revealed by those past records that also include the purchasing records, the asymmetric information could be mitigated.

The above outcomes are consistent with two of our hypotheses. First, controlling past records could mitigate the severity of asymmetric information. Second, the past purchasing records could control the asymmetric information regarding adverse selection but not the asymmetric information regarding moral hazard. Hence, after the past purchasing records are controlled, the β_2 coefficient becomes small, but it is still significant since both adverse selection and moral hazard co-exist in the market.

To further support our assertions, we conduct another Probit regression to study the relationship between the purchasing decisions for this year, i.e., choice of coverage, and past records. The outcomes are provided in Table 5. We can observe from Table 5 that the coefficient of clm_b1 is significantly negative which implies that the insured who had previously filed a claim are less likely to purchase high coverage in the current year. Furthermore, the coefficient of *clm_b1*coverage_b1* is also significantly negative which implies that the insured who chose high coverage and filed a claim in the previous year will be more likely to change to low coverage this year. The empirical evidence seems to support the view that the moral hazard exists in the market and provides indirect evidence to explain why past purchasing records could not fully eliminate the asymmetric information problems in the market.

Conclusions

Asymmetric information problems in the insurance market could arise from adverse selection, moral hazard, advantageous selection, or a combination of them. Hence, the evidence on asymmetric information that is based on different empirical studies is inconsistent. This paper investigates the asymmetric information problems for the comprehensive insurance market in Taiwan's automobile insurance industry where we have a unique dataset of a tracking sample with 216,942 observations.

The first contribution of this paper is that we propose a new method to test adverse selection against advantageous selection and to identify whether valuable information is contained in past purchasing records. We find that the insured who have previously chosen a high coverage contract tend to have more accidents in the following year. This evidence fits the prediction of Rothschild and Stiglitz (1976) and supports the existence of adverse selection.

The second contribution of this paper is that we find that the asymmetric information could be mitigated by obtaining valuable information from past purchasing and claim records for those who continue to be insured. We find that controlling the past records could decrease the β_2 coefficient of Probit regression (2), i.e., mitigate the positive conditional correlation between risk and coverage. However, there still exists significant positive conditional correlation. This is because the past records can only effectively reveal risk type information and mitigate the forces of adverse selection. Hence, for ongoing contracts, the insurance company could observe the risk type according to the previous choice of coverage of the insured since they could be screened by the menu of contracts that the insurance company has offered. Controlling the past records of coverage choice as well as claims could help to control the risk type of the insured and mitigate the adverse selection in the market. As for the forces of moral hazard, there is no compelling reason why past records could help reveal this information and also control it.

Finally, the third contribution of the paper is that we provide empirical evidence to support the existence of moral hazard in the market. The same evidence also indirectly supports our assertion that past purchasing records can not fully control asymmetric information problems with both moral hazard and adverse selection.

References

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

Cardon, J.H., & Hendel, I. 1988. Asymmetric Information in Health Insurance: Evidence from the National Health Expenditure Survey. ***Mimeo***. Princeton University.

Cawley, J., & Philipson, T.J. 1999. An empirical examination of information barriers to trade in insurance. ***American Economic Review***. 89(4): 827-846.

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

Chiappori, P.A., & Salanie, B. 2000. Testing for asymmetric information in insurance markets. ***Journal of Political Economy***. 108(1): 56-78.

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

Davidoff, T., & Welke, G. 2004. Selection and moral hazard in the reverse mortgage market. ***Working paper***.

De Meza, D., & Webb, D. C. 2001. Advantageous selection in insurance markets. ***Rand Journal of Economics***. 32(2): 249-262.

Dionne, G., Gourieroux, C., & Vanasse, C. 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. ***Journal of Political Economy***. 109(2): 444-455.

Dionne, G., Gourieroux, C., & Vanasse, C. 2003. The informational content of household decisions with applications to insurance under adverse selection, ***Working Paper***. Venice International University, San Servolo.

Dionne, G., Michaud, P., & Dahchour, M. 2006. Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France. ***Working Paper***.

Dionne, G., & Vanasse, C. 1989. A generalization of actuarial automobile insurance rating models: The negative binomial distribution with a regression component. ***Astin Bulletin***. 19: 199-212

Dionne, G., & Vanasse, C. 1992. Automobile insurance ratemaking in the presence of asymmetrical information. ***Journal of Applied Econometrics***. 7(2): 149-15.

Fang, H., Keane, M. & Silverman, D. 2006. Sources of advantageous selection: Evidence from the Medigap insurance market. ***Working paper***.

Finkelstein, A., & Poterba, J. 2002. Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market. ***NBER working paper 8045***.

Finkelstein, A., & McGarry, K. 2006. Multiple dimensions of private information: Evidence from the long-term care insurance market. ***American Economic Review***. 96(4): 938-958.

Gronqvist, E. 2004. Does adverse selection matter? Evidence from a natural experiment. ***SSE/EFI Working Paper Series in Economics and Finance No. 575.***

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.

Huang, R. J., Tzeng, L. Y. & Wang, K. C. 2008. The co-existence of adverse and advantageous selection: Evidence from automobile liability insurance in Taiwan. ***Working paper***.

Lemaire, Jean. 1989. A comparative analysis of most European and Japanese bonus-malus systems. ***Journal of Risk and Insurance***. 55(4): 660-681.

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

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., & Stiglitz, J.E. 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. ***Quarterly Journal of Economics***. 90(4): 629-649.

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

Shavell, S. 1979. On moral hazard and insurance. ***Quarterly Journal of Economics***. 93(4): 541-562.

Wang, K. C. 2006a. The asymmetric information problems in the automobile comprehensive insurance contracts of Taiwan. ***NTU Management Review***. 16(2): 161-185.

Wang, K. C. 2006b. The incentive effect in the comprehensive automobile insurance of Taiwan. ***NTU Management Review***. 17(1): 31-58.

Wang, J. L., Chung, C. F., & Tzeng, L. Y. 2008. An empirical analysis of the effects of increasing deductibles on moral hazard. *Journal of Risk and Insurance*. 75(3): 551-566.

Table 1 Definitions of the Variables

Variable	Definition
claim	a dummy variable that equals 1 when the insured has previously filed a claim[16] this year, otherwise it equals 0
coverage	a dummy variable that equals 1 when an individual chooses a high coverage, i.e., a type A or type B contract, otherwise it equals 0
clmno	observed number of claims[17] this year
Past records:	
clm_b1	a dummy variable that equals 1 when the insured has claimed during the previous year, otherwise it equals 0.
coverage_b1	a dummy variable that equals 1 when the insured chooses a high coverage contract in the last year, otherwise it equals 0.
ER1	a dummy variable that equals 1 when the bonus-malus coefficient equals or is under 0.9, otherwise it equals 0.
ER2	a dummy variable that equals 1 when the bonus-malus coefficient equals or is under 1, but above 0.9, otherwise it equals 0[18].
Demographic variables:	
carage1	a dummy variable that equals 1 when the car is one year old, otherwise it equals 0.
carage2	a dummy variable that equals 1 when the car is two years old, otherwise it equals 0
carage3	a dummy variable that equals 1 when the car is three years old, otherwise it equals 0
carage4	a dummy variable that equals 1 when the car is four years old, otherwise it equals 0
carage5	a dummy variable that equals 1 when the car is five years old, otherwise it equals 0
carage6	a dummy variable that equals 1 when the car is six years old, otherwise it equals 0
carage7	a dummy variable that equals 1 when the car is seven years old, otherwise it equals 0[19]
catpcd_1	a dummy variable that equals 1 when the car is a sedan and is for non-commercial or for long-term rental purposes, otherwise it equals 0
catpcd_2	a dummy variable that equals 1 when the car is a small freight-truck and is for non-commercial purposes or for business use, otherwise it equals 0
tramak_i	*i=n,f,h,t,c*, a dummy variable that equals 1when the trademark of the car is the assigned brand, otherwise it equals 0
channel_i	a dummy variable that equals 1 when the policy is sold through the channel i, where i=D, T, L, F, A, otherwise it equals 0.

16 When we compare type A and B contracts with a type C contract, the definition of a claim encompasses those claims from a collision, and the adjusted deductible by defining a claim for a type C contract is NT$3,000 higher than a claim for a type A or type B contract..

17 The definition of a claim is the same as it is in the previous note.

18 The smaller value of the bonus-malus coefficient means that the driver had a better record. The reference group for the two dummy variables of the bonus-malus coefficients includes those coefficients above 1.

19 The reference group for those dummy variables for the car age is that which includes all the cars used over seven years.

city	a dummy variable that equals 1 when the owner of the car lives in a city, otherwise it equals 0
north	a dummy variable that equals 1 when the car is registered in the north of Taiwan, otherwise it equals 0
south	a dummy variable that equals 1 when the car is registered in the south of Taiwan, otherwise it equals 0
east	a dummy variable that equals 1 when the car is registered in the east of Taiwan, otherwise it equals 0[20]
sexf	a dummy variable that equals 1 when the owner of the car is female, otherwise it equals 0
married	a dummy variable that equals 1 when the owner of car is married, otherwise it equals 0
age1	a dummy variable that equals 1 when the insured is between the ages of 30 and 25, otherwise it equals 0
age2	a dummy variable that equals 1 when the insured is between the ages of 60 and 30, otherwise it equals 0
age3	a dummy variable that equals 1 when the insured is over the age of 60, otherwise it equals 0[21]
Control variables:	
lnprem	logarithm of the premium for the year of each policy.
year_2002	a dummy variable that equals 1 when the data belong to the year 2002, otherwise it equals 0.
year_2003	a dummy variable that equals 1 when the data belong to the year 2003, otherwise it equals 0.
year_2004	a dummy variable that equals 1 when the data belong to the year 2004, otherwise it equals 0.
year_2005	a dummy variable that equals 1 when the data belong to the year 2005, otherwise it equals 0.[22]

[20] The reference group for those three dummy variables of area is the cars registered in the middle of Taiwan.

[21] The reference group for those dummy variables of age includes the insured who are under 25 years old.

[22] The reference group for those dummy variables of year is the year 2006.

Table 2 The Basic Statistics

Variable	Mean	Std Dev
claim	0.2439	0.4294
coverage	0.4368	0.4960
clmno	0.2949	0.5732
clm_b1	0.2439	0.4294
coverage_b1	0.5093	0.4999
ER1	0.6454	0.4784
ER2	0.3181	0.4657
carage1	0.1836	0.3872
carage2	0.1519	0.3589
carage3	0.1305	0.3369
carage4	0.1053	0.3069
carage5	0.0814	0.2735
carage6	0.0577	0.2332
carage7	0.0367	0.1881
catpcd_1	0.9770	0.1498
catpcd_2	0.0145	0.1197
tramak_n	0.0067	0.0815
tramak_f	0.0996	0.2994
tramak_h	0.0716	0.2579
tramak_t	0.3602	0.4801
tramak_c	0.0811	0.2730
channel_D	0.3814	0.4857
channel_T	0.0044	0.0662
channel_L	0.0176	0.1315
channel_F	0.0131	0.1137
channel_A	0.0209	0.1430
city	0.5815	0.4933
north	0.5479	0.4977
south	0.2284	0.4198
east	0.0258	0.1584
sexf	0.6715	0.4697
married	0.9232	0.2662
age1	0.0369	0.1886
age2	0.9134	0.2813
age3	0.0477	0.2131

lnprem	9.1547	0.8218
year_2002	0.1400	0.3470
year_2003	0.1756	0.3804
year_2004	0.1951	0.3962
year_2005	0.2268	0.4188
Number of observations	216942	

Table 3 The regression for the negative binomial model, regressed on the number of claims

Dependent Variable: *clmno* Estimated Equation (1)		
Variable	**Coefficient**	**p-value**
Intercept	-1.3149***	<.0001
clm_b1	0.1939***	<.0001
coverage_b1	0.4793***	<.0001
clm_b1*coverage_b1	-0.0173**	0.0181
ER1	-0.3041***	<.0001
ER2	-0.1325***	<.0001
carage1	0.1061***	<.0001
carage2	0.0305**	0.0335
carage3	-0.0041	0.7826
carage4	-0.0179	0.2395
carage5	-0.0249	0.1162
carage6	-0.0296*	0.0803
carage7	-0.0147	0.4254
catpcd_1	0.0619*	0.0971
catpcd_2	-0.1775***	0.0001
tramak_n	0.0059	0.8732
tramak_f	0.0005	0.9613
tramak_h	-0.0687***	<.0001
tramak_t	-0.0095	0.1765
tramak_c	-0.0328***	0.0096
channel_D	0.1423***	<.0001
channel_T	0.0677	0.1280
channel_L	0.1840***	<.0001
channel_F	0.0458*	0.0905
channel_A	-0.0627***	0.0070
city	-0.0177***	0.0054
north	-0.0936***	<.0001
south	-0.0174**	0.0532
east	0.0116	0.5508
sexf	0.1778***	<.0001
marria_	0.0246**	0.0368
age1	-0.1295**	0.0384

age2	0.0477	0.4616
age3	-0.0069	0.9148
year_2002	0.3329***	<.0001
year_2003	0.2524***	<.0001
year_2004	0.2228***	<.0001
year_2005	0.1961***	<.0001

Note: The 99% significance level is denoted by ***.
The 95% significance level is denoted by **.
The 90% significance level is denoted by *.

Table 4 Probit on coverage choice (the second stage of the two-stage method)

	Dependent Variable: $coverage$ Estimated Equation (2)			
	Model 1		Model 2	
Variables	Coefficient	p-value	Coefficient	p-value
Intercept	-22.7742***	<.0001	-12.3820***	<.0001
clmno	0.7606***	<.0001	0.4283***	<.0001
$cl\hat{m}no$	0.0283***	<.0001	1.6666***	<.0001
carage1	0.1520***	<.0001	-0.8329***	0.0001
carage2	0.6002***	0.0073	-0.1985***	0.0023
carage3	0.7334*	0.0816	0.0477**	0.0265
carage4	0.7963	0.2589	0.1590	0.4171
carage5	0.8053	0.6641	0.2265	0.6419
carage6	0.7680	0.3471	0.1280	0.7795
carage7	0.7159	0.3715	0.3572	0.6541
catpcd_1	-0.1227***	0.0030	2.5183***	0.0021
catpcd_2	2.1171	0.1742	-0.5267	0.3478
tramak_n	-0.5811***	0.0035	0.2290**	0.0124
tramak_f	0.3421***	0.0033	0.5469**	0.0311
tramak_h	0.3855***	<.0001	0.1745***	<.0001
tramak_t	0.2519***	<.0001	0.2031***	<.0001
tramak_c	0.0404**	0.0117	0.0332**	0.0335
channel_D	-0.0915***	<.0001	-0.7435***	<.0001
channel_T	-0.1608***	0.0036	-0.4416	0.1913
channel_L	-0.0346	0.2982	-0.8376	0.3961
channel_F	-0.2394***	<.0001	-0.4593*	0.0882
channel_A	0.1346***	<.0001	0.3661***	<.0001
city	0.0194**	0.0158	1.6365***	<.0001
north	0.0067***	<.0001	0.3449***	<.0001
south	-0.0108	0.3594	0.0317	0.2290
east	-0.0257	0.2948	-0.1526	0.3964
sexf	0.5812***	<.0001	-0.0842***	<.0001
marria_	-0.0526***	0.0004	-0.4865***	0.0001
age1	1.2696**	0.0221	1.9173**	0.0201
age2	1.8808	0.3578	1.7198	0.3492
age3	1.6636	0.5412	-1.4890	0.5217

lnprem	2.1157***	<.0001	1.4708***	<.0001
year_2002	0.3603***	<.0001	-0.7968***	<.0001
year_2003	0.2441***	<.0001	-0.6686***	<.0001
year_2004	0.1394***	<.0001	-0.6462***	<.0001
year_2005	0.0394**	0.0212	-0.6725***	<.0001

Note: (1) Model 1 hasn't controlled the previous year's past records in the first stage

Model 2 has controlled the previous year's past records in the first stage

(2) The 99% significance level is denoted by ***.

The 95% significance level is denoted by **.

The 90% significance level is denoted by *.

Table 5 The Probit regression of purchasing decisions this year, regressed on past records

	Dependent Variable: $\mathrm{cov}\,erage$ Estimated Equation (3)	
Variable	**Coefficient**	**p-value**
Intercept	-22.3039***	<.0001
clm_b1	-0.6687***	<.0001
coverage_b1	3.2247***	<.0001
clm_b1*coverage_b1	-0.4875***	<.0001
ER1	-0.0378	0.4570
ER2	-0.0776**	0.0394
carage1	-1.1993***	<.0001
carage2	-0.7286***	<.0001
carage3	-0.1948***	<.0001
carage4	0.0019	0.9477
carage5	0.1024	0.9288
carage6	0.1191	0.8733
carage7	0.0797	0.6287
catpcd_1	-0.0814	0.1637
catpcd_2	1.8637	0.4723
tramak_n	-0.6189***	<.0001
tramak_f	0.3319***	<.0001
tramak_h	0.2826***	<.0001
tramak_t	0.1503***	<.0001
tramak_c	0.1023***	<.0001
channel_D	-0.1164***	<.0001
channel_T	-0.3123***	<.0001
channel_L	-0.1616***	<.0001
channel_F	-0.3390***	<.0001
channel_A	-0.0297	0.4356
city	0.0918***	<.0001
north	0.1820***	<.0001
south	0.0445***	0.0060
east	0.0960***	0.0045
sexf	0.4843***	<.0001
marria_	-0.0595***	0.0029

age1	0.7965***	<.0001
age2	1.1607	0.1354
age3	0.9033	0.3215
lnprem	2.0717***	<.0001
year_2002	0.4476***	<.0001
year_2003	0.2677***	<.0001
year_2004	0.1966***	<.0001
year_2005	0.0624***	<.0001

Note: The 99% significance level is denoted by ***.
The 95% significance level is denoted by **.
The 90% significance level is denoted by *.