Decomposing Automobile Insurance Policy Buying Behavior— Evidence of Adverse Selection

Chu-Shiu Li[a], Chwen-Chi Liu[b], and Jia-Hsing Yeh[c]

Abstract

This paper examines relationships between voluntary automobile insurance policy choice and risk classification in a heavily regulated market. We investigate the existence of self-selected risk classification based on a unique dynamic data set and test whether incurred claims in one year, a proxy of risk type, is informative about policy choice in the following year. By decomposing policyholders into high and low coverage and high and low deductible groups based on policy choice in the first year, we find evidence of an asymmetric information effect in the second year for low coverage, high deductible groups but not for high coverage, low deductible groups. Our results imply the coexistence of adverse selection and other behavior, including advantageous selection and habit persistence.

Keywords: adverse selection, automobile insurance, deductible, habit persistence

[a] Department of Economics, Feng Chia University, Taiwan. E-mail: csli@fcu.edu.tw.
[b] Department of Insurance, Feng Chia University, Taiwan. E-mail: liuc@fcu.edu.tw.
[c] Department of Finance, Chinese University of Hong Kong, Hong Kong. E-mail: jasonyeh@baf.msmail.cuhk.edu.hk.

I. Introduction

Empirical tests of the existence of adverse selection have been extensive in the insurance literature. Inspired by the seminal paper of Rothschild and Stiglitz (1976), many studies have examined the well-known hypothesis of a positive correlation between risk and coverage, but most have failed to find evidence of adverse selection in markets for automobile insurance (Richaudeau, 1999; Chiappori and Salanie, 2000; Dionne et al., 2001; Saito, 2006), life insurance (Cawley and Philipson, 1999), and medical insurance (Cardon and Hendel, 2001). As a single exception, by focusing learning behavior Cohen (2005) finds a positive risk-coverage correlation for automobile insurance policyholders with at least three years of driving experience. It general, however, it seems adverse selection is not as common as was expected.

To explain the inconsistency between theory and empirical evidence of adverse selection, Chiappori et al. (2006) claim that most empirical studies were done under "strong and empirically implausible assumptions." Our paper provides an alternative view by arguing that adverse selection might not characterize the entire insurance market; rather, it may exist only for specific risk groups that are typically hidden in aggregated data sets. We examine dynamic data on multi-period contracts in a highly regulated automobile insurance market, trace individual behavior in two consecutive policy years, and decompose observations into groups with different patterns of insurance policy buying behavior. We find evidence of both adverse selection and other behavior, including advantageous selection and habit persistence.

The policies analyzed in this study are repeated contracts in the voluntary automobile insurance market in Taiwan. The contract form is similar to one mentioned in Cooper and Hayes (1987) but without commitment or renegotiation between insurers and policyholders. Consumers are free to choose between high and low coverage policies. Due to heterogeneity of policyholders, many kinds of behavior are expected. For example, some high-risk insured might exhibit adverse selection and buy more coverage while others might choose less coverage. Similarly, low-risk insured might select high or low coverage, thereby exhibiting advantageous selection or adverse selection respectively. Conditional on the choice of automobile insurance contracts in the first year, policyholders may select the second-year policy based on whether or not claims were incurred in the first year when an experience rating system is adopted. Allowing all situations, we partition the full sample into eight quasi-homogeneous groups corresponding to distinct definitions of coverage (in terms of deductibles and damages covered). Each grouping corresponds to specific insurance preferences.

We provide direct evidence showing that, with the same deductibles and conditional on choosing a low coverage policy in the first year, policyholders with no claims tended to stay

with a low coverage policy, consistent with adverse selection theory. However, for those who chose a high coverage policy in the first year, incurring no claims did not provide a strong incentive to switch to low coverage in the second year. Instead, policyholders with high coverage in the first year tended to again choose high coverage in the second year. After excluding the possible confounding effect of a new car on contract choice, habit persistence seems to be the most important determinant of policy choice. For the same level of vehicle damage coverage, low-risk individuals with a high deductible policy in the first year and no claims tended to select a high deductible policy in the second year. However, of those who chose low deductibles initially, the impact of no claims on the choice of policy in the second year is not significant.

Our findings are different from those reported in early studies (Chiappori and Salanié, 1997, 2000; Dionne et al., 2001; Richaudeau, 1999) but are consistent with more recent studies (Cohen, 2005; Israel, 2005) which find positive evidence of asymmetric information in the sense that learning and experience might chiefly determine insurance contract choices. Although the data used in this paper do not date back as far as that used in recent studies, observing switching behavior between two years permits new insight into policyholder learning channels. Disregarding possible learning in the first year, estimation using the whole sample spuriously suggests the existence of adverse selection; that is, policy choice in the current year is significantly associated with whether there are incurred claims in the previous year. However, using claims in the first year as a learning factor for the next year allows testing of the adverse selection hypothesis. By decomposing the data into high and low coverage and high and low deductible groups based on policy choice in the first year, we find evidence of an asymmetric information effect in the second year for low coverage, high deductible groups but not for high coverage, low deductible groups.

In the next section, we briefly introduce the data decomposition procedure. Empirical estimation is conducted in Section III; this includes descriptions of methodology, data, and estimation results. Section IV concludes.

II. Decomposing automobile insurance buying patterns

Two important presumptions underlying the hypothesized positive correlation between risk and coverage proposed in Rothschild and Stiglitz (1976) are that the agents are identical or “observational identical” and that decision makers know their risk class (Chiappori and Salanie, 2000). When cross-sectional data are used, it is not easy to classify agents into homogeneous groups except by demographic characteristics. In the automobile insurance market, one way to solve this problem is to focus on driving experience, as is done in Chiappori and Salanie (2000) and Cohen (2005). In this paper, the availability of a unique dynamic data set enables us to separate agents in an alternate way that might be more robust.

The data studied are two years of complete transaction records in Taiwan's voluntary automobile insurance market. There is a standard bonus-malus system in the Taiwan automobile insurance market, and the insurance authority has adopted a universal pricing formula for all insurers. More importantly, contrary to the regulation environment described in Chiappori and Salanie (2000) and Cohen (2005), the claims history of an insured is shared among all insurers by law. Although the insurance policy is a kind of repeated contract, there is no commitment or renegotiation between policyholders and insurers. Policyholders might decide to exit the market without purchasing any contract for the second year. Therefore, we focus analysis on those insured that are covered for the entire two-year period. Since insurance policies can be sold any day of the year, two full calendar years of transaction records allow us to link the contract choice and claims records in the first policy-year with contract choice in the second policy-year.

Based on contract choices in the two years (and) and on claims behavior in the first year, we construct a decision tree classifying policyholders into eight groups with similar behavior and preferences, as shown in Figure 1.

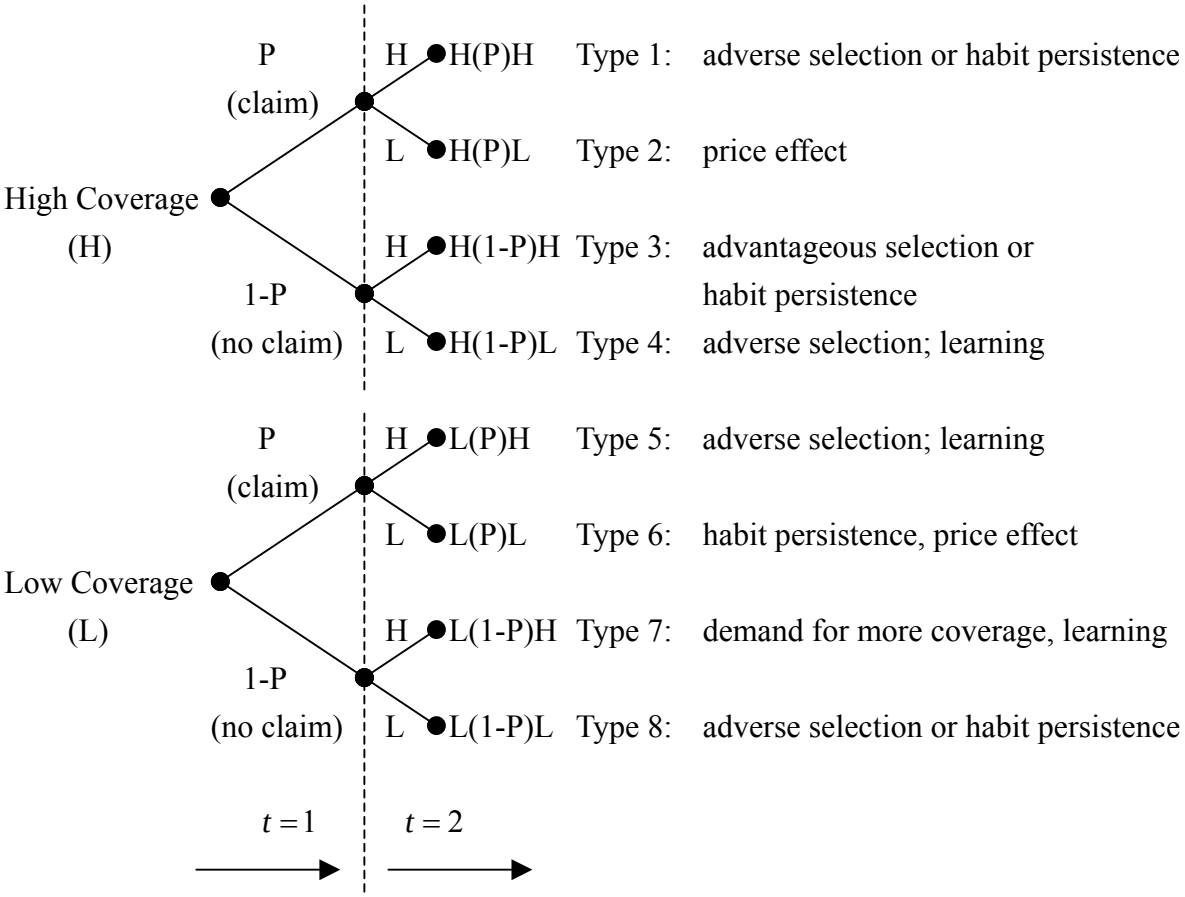

Figure 1. Decision tree of policyholders and type classifications

Adverse selection may exist if an insured chooses high coverage (H) in the first year,

claims are incurred (P) during the first year, and he or she selects high coverage (H) again in the second year, in which case the policyholder is called HPH or Type 1. Similarly, the H(1-P)H or Type 3 insured renews a high coverage policy in the second year despite incurring no claims in the first year, indicating the possibilities of advantageous selection and habit persistence. Similarly, adverse selection may occur for Types 4, 5, and 8 policyholders, whereas Type 2 behavior might be influenced by a price effect due to the experience rating system, prompting switching to low coverage even after having claims in the first period. It is worth noting that the learning effect, in terms of claims experience instead of driving experience as used in Cohen (2005) and Dionne et al. (2006), is taken into account for the contracts studied here. An insured may not know his or her true risk type initially. The choice in the first year provides an opportunity for learning. Depending on whether or not claims are incurred before the end of first year, selecting the second year contract may reveal the policyholder's true risk type.

In sum, direct testing of adverse selection can be performed by examining the evidence of a positive relationship between risk in the first year and coverage in the second year for Types 1, 4, 5, and 8 conditional on the choice of high and low coverage in the first year.

III. Empirical Analysis

(i) Data

The data used in this study are unique. We obtained data on all private automobile damage insurance contracts in Taiwan for 2002 and 2003. As in a heavily regulated industry, all insurers operate under the same operating rules and comply with obligations to submit monthly reports on every automobile insurance transaction to the Taiwan Insurance Institute, a semi-official organization responsible for publishing insurance statistics and financial data on insurers. There are three major types of coverage for vehicle damage: comprehensive form A, comprehensive form B, and moving collision coverage. As shown in Table 1, the comprehensive form A policy, sold with compulsory increasing per-claim deductibles, covers all perils. Comprehensive form B policy, sold with increasing per-claim deductibles or zero deductible, covers the same risks as form A but excludes vandalism and unknown perils. The moving collision policy, sold with no deductibles, covers only the one stated peril. Basic premiums corresponding to the coverage policies are shown in the last row. As expected, greater coverage (or no deductible) requires higher premiums.

[Insert Table 1]

In 2002, there were 509,216 vehicle damage policies with 280,485 claims; in 2003, there were 435,378 policies. We matched policy numbers in 2002 and 2003 to obtain a dynamic data set and constructed two data subsets by selecting two comparable contracts in each.

The first subset, shown in Table 2, describes those insured who chose comprehensive form B ("high coverage" hereafter) or moving collision ("low coverage" hereafter) policies in 2003[1] conditional on choices in 2002. Both contracts are sold without deductibles, but the perils covered are very different. Since the low coverage contract provides considerably less insurance protection, an examination of the relationship between incurred claims and real coverage allows for a direct test of self-selected risk classification. Under adverse selection, it is expected that a high-risk insured will tend to choose a high coverage policy and a low-risk insured will tend to choose a low coverage policy. There were 38,014 and 37,366 high and low coverage policies in 2002.

The extent of insurance coverage can be defined in many ways. In addition to perils covered, deductible choices can serve as measures of coverage. The second subset of data, shown in Table 3, summarizes policyholders who chose increasing per-claim deductibles ("high deductible" hereafter) or zero deductibles ("low deductible" hereafter) associated with comprehensive form B. Since the perils covered by these contracts are the same but deductible choices (and thus premiums) are different, the correlation between incurred claims and deductible choice can be explored. Under adverse selection, it is expected that a high-risk insured will tend to choose a low deductible policy and a low-risk insured will tend to choose a high deductible policy. The sample in this subset consists of 13,671 high deductible policies and 33,938 low deductible policies.[2]

[Insert Table 2]

[Insert Table 3]

An interesting feature of Table 2 is that the claim rate for policyholders with high coverage in 2002 and low coverage in 2003 is significantly higher than the rate for those who kept high coverage in 2003. This seems to argue against adverse selection. The claim rates given in this table create the impression that low coverage drivers have a lot fewer claims than high coverage drivers, but since the perils covered are very different, this is a misleading comparison. When it comes to the selection of deductible instead of coverage, the story remains the same. We note that premiums might play a more prominent role in the decision process because the difference between high and low coverage basic premiums (NTD 13,515) is around six times larger than that between low and high deductibles (NTD 2,314). In Table 3 it is interesting to see that, no matter what choices were made in 2002 and 2003, about one in three policyholders incurred claims in 2002. Whether or not claims were incurred in 2002 seems to have had no effect on the choice of deductible. However, the difference in perils covered between high and low coverage policies might make these ratios incomparable.

[1] The comprehensive form A policy was chosen by only 12.92% of insured in 2002 and 4.06% in 2003 and therefore was ignored for simplicity.

[2] Higher deductible options are also available. Since those who choose higher deductible options were very few, we focus on compulsory increasing per-claim deductibles of 3/5/7 for simplicity.

(ii) Dynamic testing

The hypotheses we wish to test are easy to state: Is there a positive correlation between coverage and risk? Is there a negative correlation between deductible and risk? To examine whether the choice of insurance policy contains the information about self-selected risk types, we use dynamic testing.

There are extensive discussions in the literature on empirical problems in identifying adverse selection and moral hazard when cross-sectional data is studied. Dynamic data permit us to observe the insurance buying behavior in consecutive years and may mitigate the identification bias (Chiappori, 2000).

It is clear that the choice of insurance policy cannot be determined only by claims in the previous year. In addition to risk aversion and demographic characteristics, an insured may not know his or her risk type or if a given policy is right for them. If so, we may need to investigate the temporal behavior of policyholders to verify the correlation between claims and insurance policy choice over the two-year period. To compile the desired dynamic data, we first separate policyholders in the first year into two subgroups based on contract choice, either by coverage or by deductible. Those with claims and those without claims in the first year are used as proxies of high- and low-risk groups respectively. Since the claims in the first year were actually incurred, it is more robust to use this information to classify risk type than alternative approaches (e.g., using gender or age). Therefore, self-selected behavior in the sense of Rothschild and Stiglitz (1976) may exist if there is evidence that an insured with lower (higher) coverage and with (without) claims in the first year switches to higher (lower) coverage in the second year. Empirically, we test whether changing insurance policies is associated with risk types conditional on choice in the first year. Alternatively, adverse selection may exist if an insured with higher (lower) coverage and with (without) claims in the first year tends to stay with higher (lower) coverage in the second year.

(iii) Model and variables

To model the dynamic data described above, we consider the general formulation $\mathbf{D} = f\left(\mathbf{C}, \mathbf{B}, \mathbf{X}\right)$, where $\mathbf{D}$ is an indicator matrix of self-selected coverage (deductible), representing contract choice in 2003, $\mathbf{C}$ is the incurred claims matrix for 2002, $\mathbf{B}$ is the claim coefficient matrix, and $\mathbf{X}$ is the matrix of exogenous variables. The propensity to choose a risk-type-specific insurance contract can be estimated with the linear model:

$$\mathbf{D}_i = \mathbf{C}_i\boldsymbol{\alpha} + \mathbf{B}_i\boldsymbol{\beta} + \mathbf{X}_i'\boldsymbol{\gamma}, \tag{1}$$

where the expected value of $\mathbf{D}_i$ is the propensity to choose a risk-type-specific insurance contract for observation i with incurred claims $\mathbf{C}$ in 2002, claim coefficient $\mathbf{B}_i$, and

exogenous variables $\mathbf{X}_i$. We also take into account potential nonlinear effects of $\mathbf{C}$ by considering the more general model (Dionne and Gangé, 2002; Dionne et al., 2001):

$$\mathbf{D}_i = \mathbf{C}_i\boldsymbol{\alpha} + \mathbf{B}_i\boldsymbol{\beta} + \mathbf{X}_i'\boldsymbol{\gamma} + E\left(\mathbf{C}_i \mid \mathbf{X}_i\right)\boldsymbol{\delta}\,, \tag{2}$$

where $E\left(\mathbf{C}_i \mid \mathbf{X}_i\right)$ is the approximated regressor of the expected value of $\mathbf{C}_i$ computed with the initial exogenous information $\mathbf{X}_i$.

The exogenous variables used in this paper are shown below.

Low_C_xx:	a dummy variable equal to 1 if a low coverage policy is chosen in year 20xx and 0 otherwise.
High_D_xx:	a dummy variable equal to 1 if a high deductible policy is chosen in year 20xx and 0 otherwise.
E(High_D_02):	the expected value of High_D_02 computed from the initial exogenous information, representing a nonlinear effect of the "habit persistence" deductible choice.
E(Low_C_02):	the expected value of Low_C_02 computed from the initial exogenous information, representing a nonlinear effect of the "habit persistence" deductible choice.
Counts:	the number of claims filed in 2002.
Severity:	log of total claim amount received by each policyholder.
NoClaim_02:	a dummy variable equal to 1 if no claim is filed in 2002 and 0 otherwise.
:	the expected value of NoClaim_02 computed from the initial exogenous information, representing a nonlinear effect.
Age:	the age of the policyholder.
Male:	a dummy variable equal to 1 if the policyholder is male and 0 if female.
Married:	a dummy variable equal to 1 if the policyholder is married and 0 otherwise.
Car_age:	the age of the insured automobile in years.
Exhaust:	cubic capacity of the automobile in units of 1,000 cc.
Clmcoef_xx:	the claim coefficient in year 20xx, an indicator of past driving record based on a bonus-malus scoring conversion formula and the table shown in the appendix.
District_*i*:	a set of dummy variables taking the value 1 if the automobile is

registered in district i, $i = 1,\ldots,23$, and 0 otherwise.

Company_i: a set of dummy variables taking the value 1 if the insurance policy is from company i, $i = 1,\ldots,16$, and 0 otherwise.

Premium_xx: automobile insurance premium paid in year 20xx.

Tables 4 and 5 present descriptive statistics of the two data subsets discussed above. For coverage selections in 2002 and 2003, Table 4 shows that 13.27% of insured with high coverage switched to low coverage while 97.97% of insured with low coverage stayed with low coverage. Reluctance to change or habit persistence seems to be simplest explanation for this behavior; however, the huge premium difference between high and low coverage might have some effect and cannot be ignored. The no claim ratios in 2002 for high and low coverage policies are 64.16% and 92.24% respectively. Replacing the indicator of no claims in 2002 with claim counts or severity reveals similar patterns.

It is worth noting that the figures in both tables do not reveal the gender distribution. The proportions of males in the coverage data subset are 25.78% and 35.99% for high and low coverage respectively, indirect testimonial to the fact that most automobile insurance policies are purchased under the name of a female family member. This phenomenon is easily explained by the difference in gender coefficients for males and females.[3] The nature of this phenomenon is confirmed with the marriage variable in Table 4, which shows that more than 90% of policyholders are married in all categories.

[Insert Table 4]

[Insert Table 5]

On average, the ages of cars listed on policies are 1.66 and 2.55 years for high and low coverage groups respectively. Two observations can be noted directly: new cars deserve higher coverage and, on average, it is common for owners of older cars to buy no voluntary insurance. The claim coefficient is a simple function of the cumulative claim points, which is the sum of no-claim points and claim points in the past three years. One no-claim point represents no claim in one year and corresponds to a coefficient of −0.2, starting with an initial value of 1.0. Table 4 shows that policyholders who chose high coverage in 2002 had a relatively lower claim coefficient (−0.10 in 2002 and −0.17 in 2003) than those with low coverage (−0.07 and −0.13), implying a negative relationship between the claim coefficient and coverage, supporting an advantageous selection argument.

Similarly, Table 5 shows that most policyholders (97.65% for low and 88.36% for high deductible contracts) stayed with the same policy for two years. Although not directly comparable due to different exclusions, the ratios of policies with no claims in 2002 for low

[3] According to official formula, for a given age, males have a higher pricing coefficient than females, with the difference ranging from 9% for older drivers to 19% for younger drivers.

and high deductible contracts are the same (66%). Males are similarly under-reported, admitting to 25.63% for low and 33.51% for high deductible contracts. On average, the older the car, the more a higher deductible is demanded. Comparing claim coefficients, those who chose a high deductible in 2002 had relatively lower claim coefficients; this feature is different from that in Table 4.

(iv) Empirical results

Before applying dynamic testing, it is natural to first consider simpler tests by fitting two logistic models regressing policy choices on whether claims were incurred in one year, the method proposed by Chiappori and Salanié (2000). The null hypothesis is that the error terms of these two regressions are not correlated. If the independency of error terms between policy choices and claims is significantly rejected, we can conclude that policy choice carries information about risk classification. Using our data subsets, two hypotheses can be tested: (i) that claims in 2002 and the choice of high coverage in 2003 are not correlated, and (ii) that claims in 2002 and the choice of high deductibles in 2003 are not correlated. It should be noted that policy choice in 2002 is not taken into account. Two values of the *W*-statistic are calculated using the data subsets: W_1= (-4847.02)^2/3032.19 = 7748.06 and W_2= (- 376.25)^2 /1961.88 = 72.16. Since both values are highly significant, the hypotheses are rejected. Without distinguishing contract choice in the previous year, these results are spuriously different from those found by Chiappori and Salanié (2000), incorrectly implying that the claims in the previous year have a significant effect on the selection of insurance policies in the current year.

Tables 6 through 9 report results of the dynamic testing approach, where we examining whether the association between choice of deductible or coverage in the current year and incurred claims and policy choice in the previous year contains information on self-selected risk classification.

Table 6 presents the results of logistic regressions where the response variable is Low_C_03, the probability that the policyholder stays with low coverage in 2003 conditional on choosing low coverage in 2002. In addition to the basic model, we fit two additional models that include variables to control for possible nonlinear effects. Exogenous variables include all relevant demographic and automobile factors. In particular, we consider the role of no claims in 2002 (NoClaim_02) as a proxy for risk types. We expect this variable to be positive if adverse selection exists. The results in Table 6 show that the estimated coefficients for NoClaim_02 are significant at the 1% level for all three models, revealing a strong positive relationship between no claims in 2002 and low coverage selection in 2003.[4] This

[4] We also fit regressions using Severity and Counts as exogenous variables. Reporting results of NoClaim_02 has the advantages of simplicity and easy interpretation. Its disadvantage is that some information is lost; for example, counts may follow a Poisson distribution and claims amounts may follow a Gamma distribution. Since

finding is consistent with adverse selection theory. Although there are 57 exogenous variables in the regression and all available information is used, to avoid the possible nonlinear effect on the choice of policy options, we take into account the expected value of NoClaim_02 in Models 2 and 3. The sign and significance level of coefficients for major exogenous variables remain the same, suggesting that the estimation results are robust.

In Model 3 in Table 6, an instrumental variable *E*(Low_C_02), the expected value of coverage choice in 2002 conditional on the initial exogenous variables, is taken into account for the "habit persistence" effect. Since a policyholder may exhibit inertia relative to the first year's decision, the existence of a nonlinear effect of the lagged response variable warrants examination. In light of the log likelihood values for the three models, Model 3 appears to provide the best fit. However, after controlling for this effect, our main conclusions still hold.

[Insert Table 6]

[Insert Table 7]

It is worth noting that the claim coefficient in 2003, Clmcoef_03, is significantly positively correlated to the choice of low coverage. This finding is inconsistent with the hypothesis of adverse selection, which predicts that a policyholder with a higher coefficient is more risky and would demand higher coverage. The direct link between claim coefficient and premium might explain the rationale behind this opposite effect. Although the high coverage policy is attractive, the basic premium (NTD 25,433) is more than double that for low coverage (only NTD 11,918). After including the claim coefficient, the premium difference is even higher. It is then natural that an insured with a high claim coefficient chooses low coverage due to its lower premium.

Although it is a key factor in deciding to purchase automobile insurance, premiums are not included in regressions as an exogenous variable. Since all the elements entering the premium calculation[5] are taken into account in the regressions, the premium itself is excluded to avoid multicollinearity.

Table 7 shows the results of logistic regressions for policyholders who switch to low coverage in 2003 conditional on choosing high coverage in 2002, empirically testing Types 1 to 4 in Figure 1. Most of the estimated coefficients across the three models have similar signs and significance levels as those in Table 6 except for the major regressor NoClaim_02, which reveals a significantly negative effect on switching contract behavior. This result strongly

---

the regression results are virtually the same and we do not model those characteristics, we opt for parsimony.

[5] Vehicle damage coverage premiums for all the policy options are calculated using the official formula:

where and is closely related to the vehicle age and type.

indicates the reluctance of the insured to choose low coverage even when no claims were incurred in the previous year. Two possible reasons, habit persistence and risk aversion, might explain this situation. To take into account the role of habit persistence, a nonlinear form of the contract choice in the previous year, *E*(Low_C_02), is included in Model 3 but the effect of no claims on switching from high to low coverage is still negative at the 1% significance level, suggesting that risk aversion might be the primary explanation. With all major exogenous variables significant and the smallest log likelihood, Model 3 again dominates.

The role of risk aversion on insurance demand is well documented in the literature; see for instance Hemenway (1990) and de Meza and Webb (2001). However, due to the lack of individual wealth data, further quantitative analysis of risk aversion is beyond the scope of this study. Using the available data, we are able to examine whether a new car might provide policyholders with a strong incentive to buy high coverage in both years, as the average insured car is only 1.6 years old in the relevant group. Replacing car age with a dummy variable of new car, we obtain Table 10 (discussed below). The coefficient of new car in Model 3 is negative but insignificant, failing to support the conjecture that new car owners are reluctant to switch to low coverage. In addition, after controlling for this new car effect, the results of our main finding concerning NoClaim_02 are unaffected. In conclusion, adverse selection is not supported for policyholders with high coverage in 2002.

The divergent results on the risk-coverage relationship in Tables 6 and 7 are illustrated by the data distribution in Figure 2. It is clear that for the insured with low coverage in the first year, around 90% had no claims and renewed low coverage (Type 8) in the next year. There is no doubt that the empirical evidence is consistent with adverse selection for this group. Conversely, for those who chose high coverage in the first year, Types 1 and 4 represent only 36.19% of the sample, which may not be enough to detect a positive relationship between risk in the first year and coverage in the second year. In particular, more than half of policyholders with high coverage in 2002 and no claims (57.35%) again chose high coverage as in Type 3. The negative findings in Table 7, reflecting rigid or risk-averse behavior and deserve further investigation, may be only an artifact of the data distribution.

In Tables 8 and 9, the response variable is the binary choice of high versus low deductible comprehensive form B insurance in 2003. Table 8 analyzes the situation in which no claims in 2002 has a positive or negative effect on the probability of staying with a high deductible in 2003 conditional on choosing a high deductible in 2002. The results show that the estimated coefficient for NoClaim_02 is significant at the 5% level, revealing a strong positive relationship between no claims in 2002 and staying with a high deductible in 2003. All other demographic and car related variables are not significant except for Car_age and Exhaust. The nonlinear effects play no role here. In particular, the variable of claim coefficient in 2003 is insignificant, indicating that the claim record of in the past three years

has no effect on determining the deductible choice given that the insured chose a high deductible in 2002. In sum, Table 8 supports the main theme in this paper: policyholders with a high deductible and no claims in 2002 tended to renew high deductible policies in 2003.

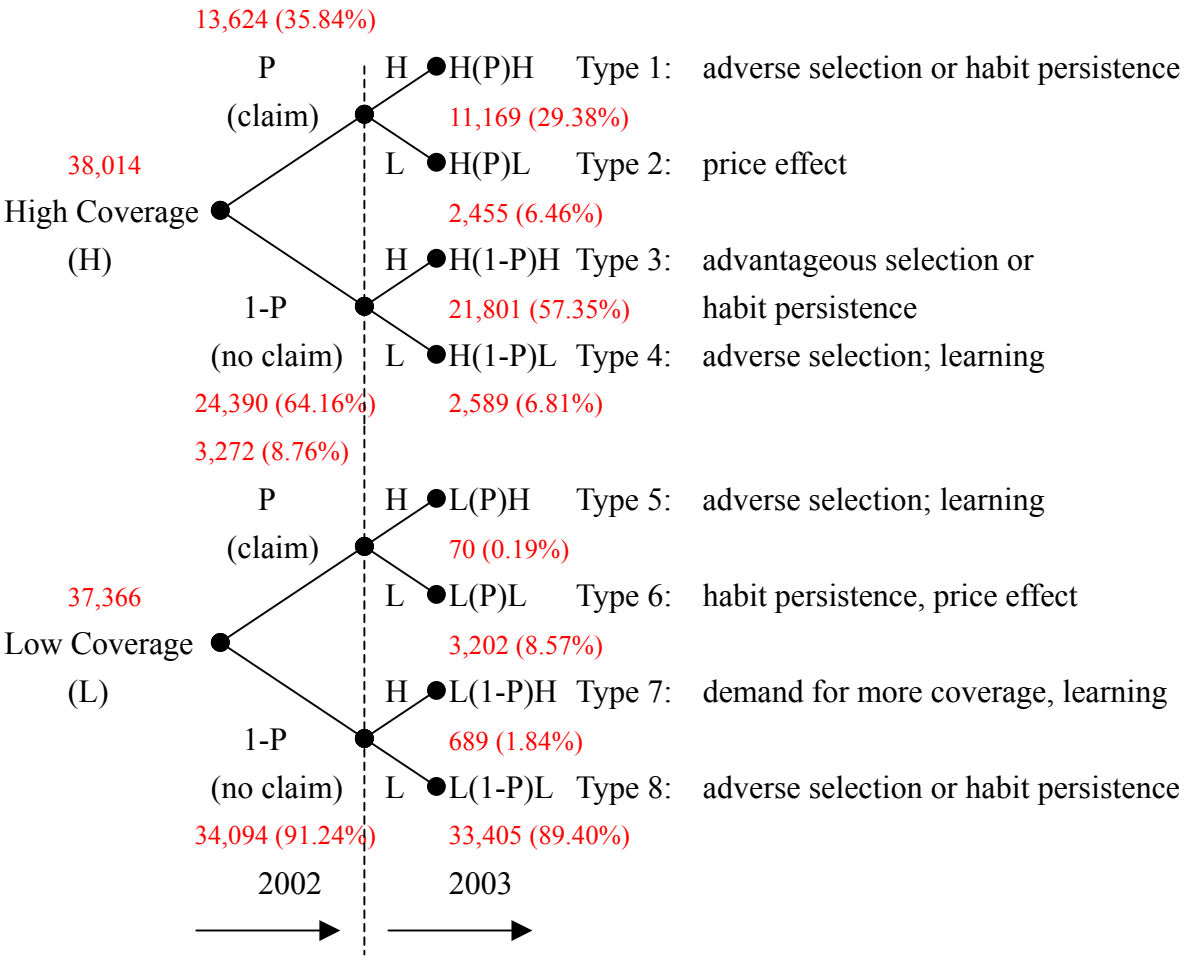

Figure 2. Data distribution of contract choice in 2002 and 2003

Table 9 presents the results of switching deductible behavior conditional on choosing a low deductible in 2002. It is interesting to see that the key exogenous variable, NoClaim_02, has a negative sign but is insignificant, indicating that having no claims in 2002 did not provide a strong incentive to switch from a low to a high deductible. After incorporating variables, *E*(NoClaim_02) and *E*(High_D_02), to control for nonlinear effects in Models 2 and 3, our main findings still hold.

[Insert Table 8]

[Insert Table 9]

In contrast with the insignificant role in the previous case, the claim coefficient in Table 9 is highly significantly positive. A high-risk insured with a larger claim coefficient may have been driven to choose a high deductible, supporting the adverse selection argument.

In Tables 8 and 9, we also take into account the role of policy choice in the first year by including an instrumental variable *E*(High_D_02). The results are opposite to those in Tables 6 and 7, as there is no evidence that the deductible choice in 2002 affect the choice of deductible in 2003. The "habit persistence" effect is strong. The log likelihood values indicate that the models in Tables 8 and 9 exhibit similar performance. In sum, nonlinear and rigid effects did not play a major role for policyholders in choosing a deductible.

In light of the results in Tables 6 to 9 where car age is considered, it is natural to inquire about the new car effect. Does having a new car play a significant role in insurance policy choice? Can it explain switching or staying behavior? Table 10 provides the results from regressions with car age replaced by a dummy variable for new car. Coefficients of this dummy variable are significant in all panels for Model 1 but are significant only in panels A and D for Model 3. In particular, for policyholders who chose low coverage in 2002, a new car tended to encourage switching to high coverage in 2003. For those who chose high coverage in 2002, there is no evidence that a new car encouraged switching or staying behavior. However, none of the coefficients of the key variable, NoClaim_02, are affected. We conclude that a new car might have some effect on policy choice but it cannot dominate the major results.

IV. Conclusion

This study examines the relationships between voluntary automobile insurance policy choice and risk classification in Taiwan's heavily regulated rating system. We test whether past claims contain information on risk types and analyze whether the policy choice is consistent with the adverse selection hypothesis in the sense of Rothschild and Stiglitz (1976).

Using two consecutive years of data, we decompose automobile insurance buying behavior into distinct groups, allowing all possible consumer preferences to coexist. The empirical evidence confirms a positive relationship between risk and coverage for those who chose low coverage in one year by showing that, after controlling a habit persistence effect, having no claims in the that year significantly attracted the insured to again choose low coverage in the next year. However, our findings concerning the claim coefficient conflict with the hypothesis since a policyholder with higher experience rating coefficient tended to stay with low coverage. The intuition behind this result may be that the premium for high coverage policy is more than double the low coverage policy. In this situation, the price (cost) effect outweighs the income (coverage) effect.

For policyholders who selected high coverage in the first year, there is no evidence of a positive relationship between risk and coverage. Instead we find that, although no claims were incurred, these policyholders preferred high coverage in both years. After accounting for policy choice inertia, risk aversion might be the main reason for this behavior, supporting

possible advantageous selection.

Comparing policies with the same coverage but different deductibles, there is evidence that choosing a high deductible is positively correlated with having no claims in the first year. However, the claim coefficient plays no role. For policyholders that chose a low deductible in the first year, we find no evidence supporting adverse selection.

In sum, a positive relationship between risk and coverage might exist for some policyholders but not necessarily for all. By exploring a unique data set, we demonstrate that it is natural for automobile insurance policyholders to exhibit heterogeneous behavior, and the data support the coexistence of adverse selection, advantageous selection, habit persistence, and a price effect as important factors determining insurance buying behavior.

References

Cardon, J. and I. Hendel (2001) "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *Rand Journal of Economics*, **32**(3): 408-427.

Cawley J. and T. Philipson (1999) "An Empirical Examination of Information Barriers to Trade in Insurance," *American Economic Review*, **89**(4): 827-846.

Chiappori, P. (2000) "Econometric Models of Insurance under Asymmetric Information," in G. Dionne, (ed.), *Handbook of Insurance*, London: Kluwer.

Chiappori, P., B. Jullien, B. Salanié, and F. Salanié (2006) "Asymmetric Information in Insurance: General Testable Implications," *Rand Journal of Economics*, Forthcoming.

Chiappori, P. and B. Salanié (1997) "Empirical Contract Theory: The Case of Insurance Data," *European Economic Review*, **41**: 943–950.

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

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

Cooper, R. and B. Hayes (1987) "Multi-Period Insurance Contracts," *International Journal of Industrial Organization*, **5**, 211-231.

de Meza, D. and D. Webb (2001) "Advantageous Selection in Insurance Markets," *Rand Journal of Economics*, 32(2): 249-262.

Dionne, G. and R. Gangé (2002) "Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance," *Journal of Risk and Uncertainty*, **24**: 213–230.

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**: 444-453.

Dionne, G., M. Pierre-Carl, and D. Maki (2006) "Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France," Working paper.

Hemenway, D. (1990) "Propitious Selection," *Quarterly Journal of Economics*, 105(4): 1063-1069.

Israel, M. (2005) "Tenure Dependence in Consumer-Firm Relationships: An Empirical Analysis of Consumer Departures from Automobile Insurance Firms," *Rand Journal of Economics*, **36**(1): 165-192.

Li, C., C. Liu, and J. Yeh (2006) "The Incentive Effects of Increasing Per-Claim Deductible Contracts in Automobile Insurance," *Journal of Risk and Insurance*, Forthcoming.

Richaudeau, D. (1999) "Automobile Insurance Contracts and Risk of Accident: An Empirical Test Using Individual Data," *Geneva Papers on Risk and Insurance Theory* **24**: 97-114.

Rothschild, M. and J.E. Stiglitz (1976) "Equilibrium in Competitive Insurance Market: The Economics of Markets with Imperfect Information," *Quarterly Journal of Economics*, **90**: 629-650.

Saito, K., (2006) "Testing for Asymmetric Information in the Automobile Insurance Market under Rate Regulation," *Journal of Risk and Insurance*, 73(2): 335-356.

Table 1. Four Types of Coverage for Property Damages to Vehicles

Insured Perils	Comprehensive Form A	Comprehensive Form B		Two-Car Collision No Deductible
		With Deductible	No Deductible	
Rollover	Yes	Yes	Yes	No
Lightening	Yes	Yes	Yes	No
Fire (Explosion)	Yes	Yes	Yes	No
Flying Objects	Yes	Yes	Yes	No
Moving Collision	Yes	Yes	Yes	Yes
Other Collision	Yes	Yes	Yes	No
Vandalism	Yes	No	No	No
Unknown Perils	Yes	No	No	No
Deductible (NTD 1,000)	3/5/7	3/5/7	0	0
Basic Premium in 2001 (NTD)	47,096	23,119	25,433	11,918

Source: The Non-Life Insurance Association, Taiwan.

Table 2. Coverage Choices in 2003 Conditional on Coverage Choices in 2002

Number of policies (Number of claims) Percent of claims		2003		Total
		High Coverage	Low Coverage	
2002	High Coverage	32,970 (11,169) 33.87%	5,044 (2,455) 48.67%	38,014 (13,624) 35.83%
	Low Coverage	759 (70) 9.22%	36,607 (3,202) 8.74%	37,366 (3,272) 8.75%
Total		33,729 (11,239) 33.32%	41,651 (5,657) 13.58%	75,380 (16,896) 22.41%

Table 3. Deductible Choices in 2003 Conditional on Deductible Choices in 2002

Number of Policies (Number of claims) Percent of claims		2003		Total
		Low Deductible	High Deductible	
2002	Low Deductible	33,141 (11,232) 33.89%	797 (301) 37.76%	33,938 (11,533) 33.98%
	High Deductible	1,591 (539) 33.87%	12,080 (4,091) 33.86%	13,671 (4,630) 33.86%
Total		34,732 (11,771) 33.89%	12,877 (4,392) 34.10%	47,609 (16,163) 33.94%

Table 4. Summary Statistics of Key Variables for the Coverage Data Subset, Coverage Choice in 2003 Conditional on Coverage Choice in 2002

	High Coverage in 2002 $N = 38,014$		Low Coverage in 2002 $N = 37,366$	
Variable	Mean	StdDev	Mean	StdDev
Low_C_02	0.0000	0.0000	1.0000	0.0000
Low_C_03	0.1327	0.3392	0.9797	0.1411
NoClaim_02	0.6416	0.4795	0.9124	0.2827
Counts	0.4452	0.6714	0.0964	0.3275
Severity	3.5634	4.7801	0.8449	2.7433
Age	42.0246	9.1678	40.9426	9.4475
Male	0.2578	0.4374	0.3599	0.4800
Married	0.9229	0.2668	0.9066	0.2910
Car_age	1.6655	1.7312	2.5532	2.0279
Exhaust	1.8942	0.6865	1.8827	0.4850
Clmcoef_02	−0.1089	0.1881	−0.0729	0.1617
Clmcoef_03	−0.1713	0.2240	−0.1314	0.2047
Premium_02	22.5190	9.8752	8.9514	4.8221
Premium_03	18.4842	9.8153	7.5578	4.5956

Table 5. Summary Statistics of Key Variables for the Deductible Data Subset, Deductible Choice in 2003 Conditional on Deductible Choice in 2002

	Low Deductible in 2002 $N = 33,938$		High Deductible in 2002 $N = 13,671$	
Variable	Mean	StdDev	Mean	StdDev
High_D_02	0.0000	0.0000	1.0000	0.0000
High_D_03	0.0235	0.1514	0.8836	0.3207
NoClaim_02	0.6602	0.4737	0.6613	0.4733
Counts	0.4128	0.6402	0.4021	0.6279
Severity	3.3810	4.7245	3.3900	4.7512
Age	42.2216	9.0841	42.1674	9.2838
Male	0.2563	0.4366	0.3351	0.4720
Married	0.9252	0.2631	0.9034	0.2954
Car_age	1.7370	1.7456	2.2400	2.0509
Exhaust	1.9028	0.7138	1.9808	0.5878
Clmcoef_02	−0.1145	0.1916	−0.1697	0.2273
Clmcoef_03	−0.1906	0.2268	−0.2448	0.2474
Premium_02	22.2400	10.0174	21.2599	11.4451
Premium_03	19.5971	9.9644	18.9843	11.2017

Table 6. Logistic Regressions on Staying with a Low Coverage Contract in 2003 Conditional on Choosing Low Coverage in 2002

Variable	Model 1	Model 2	Model 3
Intercept	22.7029*** (1.1183)	27.4365*** (0.4251)	26.5033*** (1.2872)
NoClaim_02	0.4959*** (0.1353)	0.5612*** (0.1360)	0.6401*** (0.1397)
$E\left(\text{NoClaim\_02}\right)$	—	−9.6000*** (1.3716)	−1.0265 (1.3770)
E(Low_C_02)	—	—	−−9.1830*** (0.5068)
Age	−0.0011 (0.0042)	−0.0111** (0.0044)	−0.0293*** (0.0046)
Male	0.0647 (0.0815)	0.2457*** (0.0860)	0.6827*** (0.0913)
Married	−0.0100 (0.1364)	0.0300 (0.1370)	−0.2220 (0.1403)
Car_age	0.5495*** (0.0334)	0.8769*** (0.0590)	1.3336*** (0.0654)
Exhaust	−0.0235 (0.0608)	−0.0036 (0.0620)	−0.1417** (0.0589)
Clmcoef_03	6.0575*** (0.2626)	6.5476 *** (0.2703)	9.1678 *** (0.3057)
District and Company Variables	—	—	—
Observations	37,366	37,366	37,366
Log Likelihood	−3228.3746	−3202.3346	−3026.1719

Notes: Standard errors in parentheses; *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 7. Logistic Regressions on Switching to a Low Coverage Contract in 2003 Conditional on Choosing High Coverage in 2002.

Variable	Model 1	Model 2	Model 3
Intercept	−18.0700*** (0.3353)	−16.6743*** (0.4251)	−16.5676*** (0.4312)
NoClaim_02	−0.4215*** (0.0416)	−0.4049 *** (0.0418)	−0.3470 *** (0.0422)
$E\left(\text{NoClaim\_02}\right)$	—	−2.7200*** (0.5556)	−1.4429** (0.5814)
E(Low_C_02)	—	—	−2.2603*** (0.1772)
Age	−0.0161*** (0.0018)	−0.0195*** (0.0020)	−0.0246*** (0.0020)
Male	0.1435*** (0.0370)	0.2020*** (0.0389)	0.3141*** (0.0401)
Married	−0.1315** (0.0588)	−0.1245** (0.0589)	−0.1872*** (0.0592)
Car_age	−0.0797*** (0.0114)	0.0127 (0.0218)	0.0929*** (0.0231)
Exhaust	−0.2269 *** (0.0387)	−0.2210*** (0.0387)	−0.2549*** (0.0390)
Clmcoef_03	3.3983*** (0.0849)	3.4655*** (0.0861)	4.0542*** (0.0975)
District and Company Variables	—	—	—
Observations	38,014	38,014	38,014
Log Likelihood	−13,035.4592	−13,023.6547	−12,940.7384

Notes: Standard errors in parentheses; *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 8. Logistic Regressions on Staying with a High Deductible Contract in 2003 Conditional on Choosing a High Deductible in 2002

Variable	Model 1	Model 2	Model 3
Intercept	16.8605*** (0.7916)	16.7885*** (0.8089)	16.7899*** (0.9419)
NoClaim_02	0.1594** (0.0705)	0.1578** (0.0705)	0.1610** (0.0707)
$E\left(\text{NoClaim\_02}\right)$	—	0.2450 (0.4002)	0.1512 (0.4256)
E(High_D_02)	—	—	0.7017 (1.0874)
Age	0.0002 (0.0031)	0.0003 (0.0031)	0.0004 (0.0031)
Male	0.0278 (0.0607)	0.0235 (0.0611)	0.0003 (0.0709)
Married	−0.1210 (0.1039)	−0.1240 (0.1040)	−0.1035 (0.1088)
Car_age	0.1197*** (0.0168)	0.1138*** (0.0193)	0.0923** (0.0384)
Exhaust	0.3818*** (0.0614)	0.3829*** (0.0615)	0.3591*** (0.0717)
Clmcoef_03	0.0173 (0.1432)	0.0605 (0.1597)	0.0806 (0.1627)
District and Company Variables	—	—	—
Observations	13,671	13,671	13,671
Log Likelihood	−4,484.6772	−4,484.4901	−4,484.2817

Notes: Standard errors in parentheses; *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 9. Logistic Regressions on Switching to a High Deductible Contract in 2003 Conditional on Choosing a Low Deductible in 2002

Variable	Model 1	Model 2	Model 3
Intercept	−35.1227 (2278.126)	−35.8911 (2276.778)	−35.6525 (2277.223)
NoClaim_02	−0.0707 (0.0906)	−0.0627 (0.0911)	−0.0616 (0.0911)
$E\left(\text{NoClaim\_02}\right)$	—	2.1786*** (0.5435)	2.0693*** (0.5849)
E(High_D_02)	—	—	0.6407 (1.2781)
Age	−0.0049 (0.0042)	−0.0040 (0.0042)	−0.0039 (0.0042)
Male	0.1239 (0.0809)	0.0914 (0.0814)	0.0692 (0.0927)
Married	−0.2345* (0.1339)	−0.2495* (0.1340)	−0.2323* (0.1384)
Car_age	0.0274 (0.0211)	−0.0260 (0.0258)	−0.0446 (0.0451)
Exhaust	0.0611*** (0.0222)	0.0480* (0.0224)	0.0390 (0.0288)
Clmcoef_03	0.6881*** (0.1663)	1.0284*** (0.1855)	1.0386*** (0.1864)
District and Company Variables	—	—	—
Observations	33,938	33,938	33,938
Log Likelihood	−3,465.5939	−3,457.4796	−3,457.3547

Notes: Standard errors in parentheses; *, **, and *** denote significance at 10%, 5%, and 1% levels.

Table 10. The Effect of New Car on Contract Choice

Variable	Model 1	Model 2	Model 3
Panel A: Logistic Regressions on Staying with a Low Coverage Contract in 2003 Conditional on Choosing Low Coverage in 2002			
NoClaim_02	0.5350*** (0.1360)	0.5025*** (0.1363)	0.5167*** (0.1369)
New Car	−1.8319*** (0.0966)	−1.6133*** (0.1115)	−1.8300*** (0.1165)
Log Likelihood	−3,245.8003	−3,238.7522	−3,185.1503
Panel B: Logistic Regressions on Switching to a Low Coverage Contract in 2003 Conditional on Choosing High Coverage in 2002			
NoClaim_02	−0.4312 *** (0.0416)	−0.4027 *** (0.0418)	−0.3580 *** (0.0421)
New Car	0.1274*** (0.0357)	−0.0644 (0.0428)	−0.0511 (0.0431)
Log Likelihood	−13,054.8916	−13,022.6887	−12,947.9313
Panel C: Logistic Regressions on Staying with a High Deductible Contract in 2003 Conditional on Choosing a High Deductible in 2002			
NoClaim_02	0.1791** (0.0705)	0.1664** (0.0704)	0.1719** (0.0705)
New Car	−0.2717*** (0.0707)	−0.1877** (0.0754)	−0.0087 (0.0831)
Log Likelihood	−4,504.6802	−4,499.8729	−4,487.1821
Panel D: Logistic Regressions on Switching to a High Deductible Contract in 2003 Conditional on Choosing a Low Deductible in 2002			
NoClaim_02	−0.0660 (0.0907)	−0.0646 (0.0910)	−0.0637 (0.0911)
New Car	−0.1554* (0.0835)	−0.0176 (0.0925)	−0.0455* (0.1009)
Log Likelihood	−3,464.6624	−3,457.9804	−3,457.7400

Notes: Standard errors in parentheses; *, **, and *** denote significance at 10%, 5%, and 1% levels.