Last revision: November 2001

Asymmetric Information and Learning in the Automobile Insurance Market

Alma Cohen*

Abstract

This paper tests the predictions of adverse selection models, using data from the automobile insurance market. Consistent with the presence of adverse selection, I find that low-deductible choices are correlated with more accidents, higher losses from accidents, and different distributions of the number of accidents and the damage in the event of an accident. Consistent with the presence of learning by policyholders, I find that this correlation between low deductibles and more accidents exists only for drivers who have three or more years of driving experience. Consistent with the presence of learning by insurers, I find that the magnitude of the identified correlation is smaller for repeat customers; that customers who stay with their insurer are disproportionately ones with a good claim history with the insurer; and that insurers make greater profits on repeat customers than on new customers.

JEL classification: D40, D80, D82, D83, L10, G22.
Keywords: Asymmetric information, adverse selection, moral hazard, insurance, deductible, learning of information, repeat customers.

* Department of Economics, Harvard University (acohen@kuznets.harvard.edu). I am grateful to Gary Chamberlain, David Cutler, Rajeev Dehejia, David Genesove, Liran Einav, Shigeo Hirano, Caroline Hoxby, Ariel Pakes, Daniel Paserman, Jack Porter, Manuel Trajtenberg, Richard Zeckhauser, David Wise, and participants in seminars at Harvard, Tel-Aviv University, and Hebrew University for their valuable suggestions. I also wish to thank the IDI Company for the data and Shai Fogel, its CEO, for very helpful discussions about the company and its market.

I. INTRODUCTION

Ever since the seminal works of Akerlof (1970) and Rothschild and Stiglitz (1976), economic theorists have devoted much effort to developing models of adverse selection. This paper seeks to test the predictions of adverse selection models, using data from the automobile insurance market.

The paper uses a unique and rich data set that I obtained from an insurer operating in the automobile insurance market in Israel. Because the data includes *all* the information known by the insurer about its customers, it is especially fitting for studying adverse selection. The data includes the insurer's information about more than 200,000 policies issued during a five-year period. In particular, the data contains (i) all the characteristics of policyholders known to the insurer, (ii) the price-deductible menus offered to potential customers, (iii) the deductible choices made by policyholders, (iv) the past accident history reported by each policyholder when joining the insurer, and (v) the claims and payments resulting from each policy.

Adverse selection models predict a separating equilibrium in which high-risk types purchase more insurance coverage. Consistent with this prediction, I find that policyholders with low deductibles are associated with more accidents and higher total losses from accidents. Furthermore, I find that low-deductible policyholders are characterized by a different distribution of the number of accidents and a different distribution of the damages in the event of an accident.

Adverse selection results from informational asymmetries between policyholders and insurers, which in turn might be influenced by learning over time. I find evidence consistent with learning by both policyholders and insurers. As to the learning by policyholders, I find that, for drivers with no or little experience, low deductibles are not associated with more accidents, and the data is thus inconsistent with adverse selection with respect to such policyholders. With no or little driving experience, such policyholders have had relatively little opportunity to obtain

private information about their risk type and thereby gain an informational advantage over the insurer. However, such an association does exist for policyholders with three or more years of driving experience (who constitute a very high majority of the policyholders in this market).

As to the learning by the insurer, I find evidence that the association between low deductibles and more accidents is especially strong for new customers or ones who have been with the insurer for a relatively short period, and that it diminishes and eventually disappears with time for repeat customers. Once a policyholder has stayed with the insurer for a sufficiently long period, the insurer can infer the policyholder's risk type from the policyholder's record.

It is worth putting the above findings on the association between insurance coverage and risk type in the context of current research. Two recent JPE papers by Chiappori and Salanie (2000) and Dionne, Gouriéroux and Vanasse (2001) suggest that, in the automobile insurance market, higher insurance coverage is not associated with more accidents and that the evidence for this market is inconsistent with adverse selection. This thought-provoking work has not gone unnoticed. In the materials issued recently in connection with announcing the award of the Nobel Prize for work on asymmetric information, the Swedish Royal Academy of Science (2001) cites the Chiappori - Salanie findings as a reason for viewing the evidence on the presence of asymmetric information in markets as "mixed." These findings also motivated the recent effort by De Meza and Webb (2001) to develop a model of the insurance markets (with agents whose risk-aversion is negatively correlated with their risk type) that does not predict an association between higher insurance coverage and more accidents.

In contrast to this recent empirical work, using a data set that is more complete in significant ways than the data used in the above studies, and verifying that my results hold also using the same methodologies as employed by them, I obtain robust results that are consistent with the predictions of basic adverse selection theory. Furthermore, I explain how Chiappori and Salanie

(2000) did not find such a correlation because of their limiting their testing to policyholders with 0-2 years of driving experience. While I confirm that such a correlation does not exist for policyholders with so little driving experience, I show that this finding does not at all carry over to policies sold to individuals with 3 or more years of driving experience, which make up roughly 98% of the total policies sold by the companies studied by both Chiappori-Salanie and myself.

In addition to examining whether the evidence is consistent with the presence of informational asymmetries between insurers and policyholders, I also examine whether the learning that insurers do about their customers provides them with an informational advantage with respect to these customers over competing insurers. When a customer changes insurers, the information that the new insurer would have about the customer's past (from the customer's own reporting or from inferences drawn from the customer's making a switch) would not be as full and accurate as the one that the customer's current insurer has. I find that the evidence is indeed consistent with such differences in information among insurers.

In particular, I show that the costs to the insurer of new customers are higher than would be predicted if the customers' reports of past claims were assumed to be as reliable as the insurer's records about repeat customers. Also, I find that policyholders who leave the insurer are disproportionately ones who have a poor past record and who thus could benefit from switching to an insurer that would know less about them. Policyholders who remain with the insurer, in turn, perform subsequently better than new customers. Furthermore, consistent with the possibility that insurers obtain private information and thus market power with respect to some repeat customers, I find that the insurer makes greater profits on repeat customers than on new customers.

It should be noted that, although finding correlation between low deductibles and more accidents is a necessary condition for adverse selection to be exist, it is not sufficient to *establish* the presence of adverse selection. Such a correlation can be explained not only by hidden information (adverse selection) but also by hidden action -- that is, moral hazard. Under a moral hazard story, a customer's higher risk type does not cause the choice of a low deductible, but rather the customer's higher risk type is *caused* by having a low deductible which induces the taking of low precautions.[1] Therefore, the recent empirical work that found no such correlation suggested that this market lacks not only adverse selection but also moral hazard. Thus, by themselves, my findings that such a correlation exists keep alive the possibility in this market of not only adverse selection but also moral hazard.

Because adverse selection and moral hazard both produce a correlation between coverage and more accidents, the problem of disentangling is a difficult one for the empirical study of markets with asymmetric information (Royal Academy of Science (2001)). I point out that some of my findings concerning the dynamics of behavior over time – concerning differences between young and experienced drivers, between departing and staying customers, and between new and repeat customers -- do not appear readily explainable by a standard moral hazard story. A full testing for moral hazard in this market, however, is beyond the scope of the present study and it is pursued in a separate work in progress (Cohen and Deheija (2001)).

The analysis of this paper is organized as follows. Section II describes the predictions of adverse selection models to be tested. Section III discusses prior empirical work. Section IV describes the data used for the study. Section V analyzes whether low-deductible policyholders are associated with higher risks. Section VI analyzes whether the evidence is consistent with learning by policyholders and/or insurers over time. Finally, Section VII concludes.

[1] See Spence and Zeckhauser (1971), Shavell (1979), and, for a recent review, Winter (2000).

II. Adverse Selection Predictions

A. Sorting

As is common, the insurer whose data I use offers a menu with different levels of deductibles (and associated levels of premium). Policyholders that choose different deductibles must be different. They might differ in their risk-aversion,[2] and such differences would not affect the costs of the policy to the insurer. Policyholders also might differ in the risk that they pose to the insurer. Such differences are relevant for the costs of the insurer and they might give rise to adverse selection.

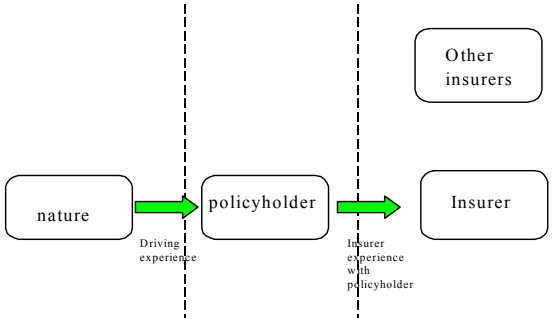
The basic theory of adverse selection was developed in the seminal work of Rothschild-Stiglitz (1976) and by the subsequent works of Riley (1979), Spence (1978), Stiglitz (1977), Miyazaki (1977), Wilson (1977), and Grossman (1979).[3] According to the standard adverse selection theory, a menu of different deductibles will produce a separation, with high-risk individuals choosing low deductibles.

B. Learning and the Policyholder-Insurer Asymmetry

Adverse selection appears when purchasers of insurance policies have an informational advantage over the insurer. Whether and to what extent such informational asymmetries exist might well depend on the learning of information by the policyholder and by the insurer. The diagram below indicates how learning can close informational gaps between the policyholder and nature and between an insurer and a policyholder.

---

[2] One reason for having deductibles (see e.g., Arrow (1963), Raviv (1979)) is to eliminate coverage for small claims, because such coverage would produce administrative costs (which are ultimately borne by policyholders) but would provide only minimal benefit in terms of saving risk-bearing costs. The greater an individual's degree of risk-aversion, the lower the optimal deductible for this individual. Since individuals vary in their degree of risk-aversion, insurers should offer a menu of policies with different deductibles, and individuals with different degrees of risk-aversion can be expected to choose different deductibles.



One type of learning is by policyholders. While nature accurately knows any individual's risk-type, the individual might well be imperfectly informed about this risk-type when getting the driving license. The more driving experience a policyholder has, the more information the policyholder might obtain about the policyholder's risk type. Thus, if driving experience improves policyholders' information about their risk type, it can be expected that, other things equal, the more driving experience a group of policyholders has, the stronger the correlation between low deductibles and higher risks.

Another type of learning might take place on the part of insurers. As the diagram indicates, the longer a policyholder stays with the same insurer, the more information the insurer will obtain about the policyholder's risk type. When a policyholder joins an insurer, the insurer will be in the same position as all other insurers in the market. If the policyholder remains with the insurer for some time, however, the insurer will be able to update the prior that it had concerning the policyholder's risk type when the policyholder joined. Thus, it can be expected that, other things being equal, the more experience an insurer has with a group of customers, the smaller the correlation if any between low deductibles and high risks.

---

[3] For excellent recent surveys of adverse selection models, see Riley (2001) (who surveys such models in general), and Dionne, Doherty, and Fombaron (2000) (who focus on such models of insurance markets).

In thinking about learning by insurers and policyholders, it is also useful to distinguish between (i) the period (if any) from a customer's getting a license until the customer's joining the studied insurer and (ii) the period from the customer's joining the insurer on. During the first period (i), the realization of risks produced by the customer's driving is directly observed only by the customer (and the customer's insurance at the time) and not by the studied insurer. Thus, since the learning in this period is done only by the customer and not by the studied insurer, the longer this period, the greater the expected asymmetry of information between the customer and the insurer when the customer joins the insurer.

In contrast, during the period in which the customer is with the studied insurer, the realization of risks is observed both by the customer and the insurer. Assuming that they both get the same observations, i.e., that the flow of new information to each one of them is symmetric, the updating done on both sides will work to reduce whatever informational asymmetry the customer had when joining the insurer. The reason is that the informational benefit from observing the realization of risks in a given year is larger for the insurer than for the customer since the customer already observed the realizations of risks in one or more years of driving before joining the studied insurer. Thus, under this assumption, it can be expected that, the longer the period that a group of policyholder has been with the insurer, the smaller the correlation if any between low deductibles and more accidents.

C. Learning and Differences in Information among Insurers

As the above diagram also indicates, when an insurer learns about the risk type of a repeat customer that has stayed with the insurer for some time, such learning about the customer's risk type might create a difference in information between the insurer and other insurers (to which the customer might subsequently turn) that would not have otherwise existed. Of course, if the repeat customer does turn to another insurer, this other insurer could ask the customer about past claim

history. There is evidence, however, that such self-reporting is often neither full nor accurate (AIRAC, 1984). There is also evidence that most of the accidents for which claims are submitted are not reported in public records (Insurance research Council (1991)), and new insurers thus cannot obtain the full past claim history of a new customer by inspecting public records. Therefore, a new insurer would likely have an informational disadvantage compared with the customer's incumbent insurer.[4]

To be sure, a new insurer could also draw inferences from the fact that a new customer has decided to switch from the customer's incumbent insurer. If all switchers were ones with a bad record, the insurer would be able to infer that a switcher must have such a record. But in a world in which some individuals change insurers for other reasons – say, relocation or unpleasant experience with the insurer's service – an insurer generally would not be able to infer fully a customer's past record from the customer's decision to switch. In particular, when a customer who in fact has a bad record switches to another insurer, this other insurer would not be able to tell for sure that the switch has been motivated by the customer's bad past record.

This adverse selection story with informational differences among insurers has predictions that I will test. First, if information about claim history obtained from reporting by new customers is not as full or accurate as information obtained about repeat customers from the company's own records, then the former will be less good a predictor of subsequent performance than the latter. Second, customers with a poor record with an insurer will have something to gain from switching to a new insurer who would have less information about their past history; accordingly, customers that leave their insurer will be disproportionately ones with a poor past record. Customers who stay with the insurer, in turn, will subsequently perform relatively well.

[4] In the market under consideration, as in most markets, there are no laws or agreements among insurers that require insurers to share with other insurers the records that they have with respect to departing customers. Whether such a requirement would be desirable is studied by Fombaron (1997b).

D. Learning and the Pricing of New and Repeat Customers

A substantial work has been done on developing multi-period models of adverse selection. Some of these models have focused on the optimal design of policies that commit customers to a multi-period contract (Dionne (1983), Dionne and Lasserre (1985, 1987), Cooper and Hayes (1987)) or that at least commit the insurer to offer the policyholder certain policies in subsequent periods (Dionne and Doherty (1994)). Although such policies are observed in certain countries (see, e.g., Dionne and Vanasse (1992)), there are many cases of policies that are for a single period only and involve no commitments on the part of either the customer or the insurer (Kunreuther and Pauly (1985)). In particular, this is the case for all the insurers operating in the Israeli automobile insurance market, including the insurer whose data I study.

For our purposes, then, the relevant models are the "no-commitment" models that were developed by Kunreuther and Pauly (1985), Fombaron (1997), and Nilssen (2000). The informational advantage that insurers obtain over competing insurers with respect to their repeat customers, which was discussed earlier, plays an important role in these models. When an insurer has better information about repeat customers' types than competing insurers, the insurer will have some market power with respect to repeat customers that it has identified as low-risk types. Thus, the insurer will be able to charge these customers more than the break-even price (given the low risk that they pose), because these customers will not be able to obtain this break-even price from other insurers to which the customers' low risk type is not known for sure.

Furthermore, these dynamic models predict *lowballing* with respect to new customers. When a new customer joins an insurer, the insurer will anticipate that it will be able to charge more than the break-even price down the road if the customer turns out to be of a low-risk type. Therefore, the insurer will be willing to charge new customers less than the break-price for the first policy sold to them. Because insurers can thus be expected to over-charge repeat customers

and to under-charge new customers, they will make higher profits, and therefore have a lower loss ratio, on repeat customers than on new customers. This prediction is also one that I will test.

III. Prior Empirical Work

The presence of adverse selection has been well documented in some insurance markets. In particular, Cutler and Zeckhauser (2000) conclude that the evidence indicates that adverse selection is present – and is quantitatively large – in the health insurance market. Evidence of adverse selection has also been found for the annuity market (Friedman and Warshawsky (1990), Bruggiavini (1993), and, most recently, Finkelstein and Poterba (2000)).[5] However, recent work has argued that the evidence is inconsistent with the presence of adverse selection in the market for automobile insurance.

It should be first noted that prior work on this market has not used individual data to study the dynamic issues of learning, choice of insured, and pricing over time that I study in the second half of this paper. No work has studied how informational asymmetries between customers and insurers might depend on learning by the customer (from its driving experience) or by the insurer (from its experience with the customer). As to the temporal pattern of profits on new and repeat customers, the two studies that considered it (D'Arcy and Doherty (1990) and Dionne and Doherty (1994)) used aggregate data, and their conclusions went in opposite directions (one suggesting lowballing with respect to new customers and one suggesting highballing with respect to such customers).

The focus of prior work on the automobile insurance market has been on testing the basic adverse selection prediction that larger insurance coverage is correlated with more accidents.

[5] Evidence consistent with the presence of adverse selection has also been found for some markets outside insurance. See, e.g., the study by Genesove (1993) on the wholesale used car market.

Three studies suggested the presence of adverse selection, but their findings were criticized by subsequent research as unreliable. The problem with the studies of Dahlby (1983) and Dahlby (1992) was that they did not have individual data on coverage. Puelz and Snow (1994) did use individual data obtained from a Georgia insurer, but subsequent work strongly questioned their results. The problem with their study was the lack of data on the size of claims. Finding that low-deductible policyholders had more claims, they inferred that such policyholders were associated with more accidents. But more claims do not necessarily imply more accidents: even if low-deductible policyholders had the same number of accidents as high-deductible policyholders, low-deductible policyholders would file more claims, as they would also make claims for accidents with losses falling between the levels of low and high deductibles. As will be explained, my data includes the size of the claims filed by policyholders and thus enables addressing this problem; in particular, it enables identifying whether the larger number of claims filed by low-deductible policyholders results merely from being able to file claims for a wider range of losses or also from having a larger number of accidents.

Dionne, Gouriéroux, and Vanasse (2001) recently raised another objection to the Puelz-Snow study, arguing that its results suffer from non-linearity bias. Addressing the problem suggested by these authors, the results that I present below are all robust with respect to this possible problem; in unreported regressions, I controlled for the suggested non-linearity bias using the method employed by these authors, and I obtained the same robust results throughout.

Chiappori and Salanie (1997, 2000) suggest that the correlation between deductible choice and risk type should be tested using bivariate probit, and Chiappori and Salanie (2000) apply this test to French data. Two kinds of insurance coverage are offered in France, and the authors tested whether individuals who bought higher coverage turned out to be riskier. Finding no such correlation, they inferred the absence of adverse selection.

It is important to note that this study limited the testing to drivers who obtained their license recently. Because the authors did not have the data that the insurer had with respect to customers' claims history, they could not control for this parameter, and therefore they opted to include in their analysis only drivers who had a short or non-existent driving history and thus could not have differed much in their claims history. This focus of the study, however, implies that it tested only the presence of adverse selection for beginning drivers. The lack of adverse selection in the case of such drivers, who have had little chance to develop an informational advantage over the insurer, does not imply that adverse selection does not exist with respect to more experienced drivers.

Because my data includes all the information concerning past claim history known to the insurer, I am able to conduct tests with respect to both beginning and experienced drivers. As will be discussed in detail later, my analysis confirms the Chiappori-Salanie finding that correlation between higher insurance coverage and more accidents does not exist for beginning drivers. The analysis indicates, however, using both the Chiappori-Salanie tests and other tests, that such correlation does exist with respect to experienced drivers and that adverse selection thus cannot be rejected for the automobile insurance market as a whole.

IV. THE DATA

A. The Company and its Market

The paper is based on data that I received from an insurance company that operates in the market for automobile insurance in Israel. The company started its operation in November 1994 and was the first and, for some time, the only company in Israel that marketed insurance to customers directly rather than through insurance agents. By the end of the studied period, the company was selling about 5-7% of the automobile insurance policies sold in Israel.

Direct insurers operate in many countries including the US, and they appear to have a significant cost advantage (see Cummins and Van Derhei (1979)). The studied company estimated that selling insurance directly saves roughly 25% of the administrative costs involved in marketing and handling policies. Despite their cost advantage, direct insurers generally have had difficulty in making inroads beyond a part of the market (see D'Arcy and Doherty (1990)). The reason for this is that, although their product offers similar coverage to that offered by conventional insurance, it also involves the "disamenity" of not having an agent to work with and to contact in case of need. The costs of this disamenity appear to be substantial for most but not all potential buyers of insurance. Consequently, only a subset of the insurance buyers are open to considering buying insurance directly and thus constitute potential customers of direct insurers.

B. Variables

The data contains information about 216,524 policies, and about 111,138 different policyholders, for the years 1995-1999. Each observation has the following variables with respect to the individuals purchasing the policy (the list of all variables appears in Appendix I): (1) *Buyer's demographic characteristics*: age, education, gender, family status, place of birth, immigration year, and place of residence; (2) *Buyer's car characteristics*: size of engine, model year, value of the car, value of the radio, commercial vehicle or not, main vehicle or not, type of protection against theft; (3) *Buyer's driving characteristics*: license years, number of claims in the past three years, using the vehicle on Saturdays or not, young driver or not, age of young driver, sex of young driver, license years of young driver, whether the policyholder had insurance in the past, additional drivers (if any); (4) *Menu of contract terms offered:* whereas other insurance companies offered only policies with a regular deductible, this company offered four deductible-premium alternatives – low-deductible, regular-deductible, high-deductible, and

very-high-deductible contracts -- which will be described in detail in subsection C below; (5) *Deductible choice*: what deductible (and accompanying premium) was chosen by each customer; (6) *Period covered:* the length of the period for which the customer purchased coverage; and (7) *Realization of risks covered by the policy*: the number of claims submitted by the policyholder and a description of each submitted claim, including the amount of damages reported and the amount that the insurer paid or was expected to pay. Table 1 displays descriptive statistics of the variables.

I also received from the company the estimate that it used in its calculations for the average administrative costs involved in processing a claim, which was equal to 450 NIS (~$110). When calculating for some regressions the company's total profits on certain customers, I included these estimated costs of processing claims in the company's total costs. Because these processing costs were only estimates and not "hard" number like the other variables, I checked whether the results of these regressions hold ignoring these costs and I found that they do.

C. The Deductible-Premium Menu

Israeli insurers are allowed to develop their own formula for determining insurance premia, though they must submit them for approval by the insurance regulator. The only factors that the regulator does not allow insurers to use in setting the premium are place of birth, place of residence, occupation, and education. The insurer under study attempted to take into account in its pricing decisions all the information that it was allowed to use.

The insurer offered its potential customers a menu of contract choices after first obtaining from them all the information described in subsection B. The potential customer was then given a menu of four premium-deductible contracts. One option, which was labeled "regular" by the company, offered a "regular" deductible and a "regular" premium. The term "regular" was used

for this option because the level of the regular deducible was relatively similar to the deductible levels offered by other insurers and because this deductible level was the one chosen by most policyholders. The regular premium was a function of all the characteristics of the policyholder according to the insurer's formula. The regular deductible was set at the level of 50% of the regular premium that was associated with the regular deductible, except that the regular deductible was capped at 1400 New Israeli Shekels (NIS) (about $350).

The three other price-deductible contracts that the menu offered to customers included were: 1) a "low" deductible, set at 60% of the level of the regular deductible, coming with a premium equal to 1.3 times the level of the regular premium; 2) a "high" deductible, set at a level equal to 1.8 times the level of the regular deductible, coming with a premium equal to 0.7 time the regular premium; and 3) a "very high" deductible, set a level equal to 2.6 times the level of the regular deductible, and coming with a premium equal to 0.685 times the regular premium.

Since I did not have an access to the company's formula, I regressed the regular premium quoted to each customer on the customer's characteristics (those that the company was allowed to use in its pricing) to test how well a linear regression can explain the premium. The regression, which appears in Table 2, has an R^2 of 0.71, which indicates that a linear model can be used instead of the actual formula.[6]

D. Deductibles and Claims: Summary Statistics

Table 3 provides summary statistics for the whole period covered by the data. The table indicates that, compared with the regular-deductible policies, the company did overall worse

[6] Therefore, in some of the regressions, I will for simplicity use the premium of the policyholders instead of their characteristics. In all regressions in this paper, I have verified that the results are robust and hold either using the characteristics of the policyholder or using the premium instead.

with respect to low-deductible policies, both in terms of claims' frequency and loss ratio,[7] and did overall better with respect to policies with high or very high deductibles.

Since only a tiny fraction of the customers chose high or very high deductibles (apparently the company did not price them low enough to make them attractive), my focus below will be on differences between low-deductible and regular-deductible policyholders. Individuals who chose "low" level of deductible had a larger incidence of one or more claims during the life of the policy. For example, 27.7% of low-deductible policyholders had at least one claim, whereas only 21.33% percentage of regular-deductible policyholders had at least one claim.

Note that low-deductible policyholders were able to file claims also for accidents whose damages were too small to claim under policies with a regular, high, or very high level of deductible. Thus, it is useful to compare low-deductible and regular-deductible policyholders in terms of the number of claims of a type that can be submitted by both groups of policyholders. The data indicates that, counting only claims for damages exceeding the level of the regular deductible (claims that could be submitted by both low-deductible and regular-deductible policyholders), the percentage of policyholders filing such claims is higher for low-deductible policyholders (22.9%) than for regular-deductible policyholders (17.7%). Similarly, counting only claims for damages exceeding 1.5 times the level of the regular deductible, I find that the percentage of policyholders filing such claims is significantly higher for low-deductible policyholders than for regular-deductible policyholders.

[7] Loss ratio is the ratio of the insurer's total payments for claims over the total premia received from the buyers of insurance. (See Appendix II for a precise specification.)

V. Are Lower Deductibles Associated with Higher Risks?

The summary statistics presented in section IV suggest that low-deductible policyholders might be associated with higher loss ratios and higher claim frequencies. This Section studies more systematically whether such an association exists. In particular, I test whether low-deductible policyholders are associated with: (a) more accidents, (b) higher total losses from accidents, and (c) different distributions of the number of accidents and/or the loss in the event of an accident. I focus on differences between low-deductible and regular-deductible policyholders because, as noted, the overwhelming majority of individuals in the data belong to these two groups, with less than 2% choosing high or very high deductibles.

A. The Number of Accidents

I start by comparing low-deductible and regular-deductible policyholders in terms of the number of claims submitted for accidents. As noted earlier, this comparison should focus on claims of a type that can be submitted by both types of policyholders. If we were to count on all claims reported by the low-deductible policyholders, then we would expect to find more claims submitted by low-deductible policyholders even if there were no difference in risk type between the two groups, because low-deductible policyholders can submit claims with respect to a larger range of accidents. Below I compare these two groups in terms of claims that exceed the level of regular deductibles and thus can be submitted by both groups. As will be noted below, the results hold also when making the comparisons in terms of claims exceeding certain higher thresholds.

I first test for an association between low deductibles and more accidents using a linear regression. For the set of all the policyholders choosing either low or regular deductible, I regressed the number of claims exceeding the regular deductible on the level of the regular

premium, which reflects the characteristics of the policyholder and the vehicle, and a dummy variable representing whether a regular or a low deductible was chosen.

The results, which are displayed in Table 4 (column 1), indicate that the number of claims exceeding the regular deductible is higher for low-deductible policyholders than for regular-deductible policyholders. Low-deductible policyholders had on average 0.03 claims more than regular-deductible policyholders (at the 1% confidence level). This difference is significant relative to the average number of claims that exceeded the regular deductible for either low- and regular-deductible policies. The average number of claims exceeding the regular deductible is 0.23 for low-deductible policyholders and 0.18 for regular-deductible policyholders (see Table 1). Low-deductible policyholders have about 20% more such claims than regular-deductible policyholders in each of the five years included in the data.[8]

The second test that I used to test for an association between low deductibles and more accidents is the Bivariate Probit used by Chiappori and Salanie (2000). The Bivariate Probit estimates the correlation ρ between the error terms of two binary equations. The two binary equations are the choice of the deductible on the policyholder's characteristics and the occurrence of at least one claim on the policyholder's characteristics. If the error terms of the two equations are independent, then ρ will be equal to 0; otherwise ρ will reflect the sign of the correlation between the two equations. The results, which are shown in Table 5, provide an estimate for ρ that is negative, statistically significant (at the 1% confidence level), and equal to –0.057. Thus, the hypothesis that the two equations are independent can be rejected. The negative correlation implies that, compared with regular-deductible policyholders, low-

[8] Regressing the same regressions with the characteristics of policyholders instead of using the premium yields the same coefficients on the regular deductible dummy.

deductible policyholders are more likely to submit one or more claims exceeding the regular deductible.

It might be suggested that regular-deductible policyholders might sometimes be reluctant to submit claims for accidents whose damage exceeds the regular deductible but just barely. They might not submit such claims, so the argument goes, in order to save the transaction costs involved in submitting a claim and/or to avoid having a claim in their record that might lead to an increase in the premium in subsequent years (see Hosios and Peters (1989)). To ensure that my results here or later in the paper are not vulnerable to this problem, I did the tests discussed above also for claims exceeding 1.5 times the level of the regular deductible as well as for claims exceeding 2 times the level of the regular deductible. I obtained in both cases a similar result, that the number of such claims is higher (and in a statistically significant way) for low-deductible policyholders than for high-deductible policyholders. It is worth noting that I followed a similar procedure also for the other tests in this paper that involve the number of claims, and obtained similar results throughout, and that all of the results reported below are thus robust to this problem.

B. Losses from Accidents

I then compared low-deductible and regular-deductible policyholders in terms of the costs to the insurer produced by claims exceeding the regular deductible. I regressed the total insurance payments made by the insurer in connection with such claims on the regular premium (reflecting the characteristics of the policyholder and the vehicle) and on a dummy variable reflecting whether a regular or low deductible was chosen.

The regressions, which are displayed in Table 4 (column 2), indicate that such total insurance payments are higher (at the 1% confidence level) for low-deductible policyholders.

This result holds for each one of the five years that the data includes. The regression indicates that the insurer's total insurance payments for accidents exceeding the regular deductible was higher for low-deductible policyholders than for regular-deductible policyholders by 230-315 NIS (~$58-$78) at the 1% confidence level. This increase is roughly equal to 20% of the average total insurance payments to regular-deductible policyholders.

C. The Distribution of Loss in the Event of an Accident

I next turn to examine whether low-deductible policyholders and regular-deductible policyholders have the same distribution of the damage in the event of a claim. In estimating these distributions, I need to take into account that the accidents included in the data are only those arising from accidents whose damage exceeds the level of the deductible. Therefore, I will assume that the distribution of reported damages is the truncated part of an underlying distribution of accidents – the reported accidents are those whose damage is above the level of deductible chosen by the insured.

To examine whether the distribution of the damage in the event of an accident is the same for low-deductible and regular-deductible policyholders, I estimated the loss distribution of the damage for both the low- and the regular-deductible policies. I considered three possible types of distributions: (1) Exponential, (2) Pareto, and (3) Weibull. Using Maximum Likelihood estimation, I found that the best approximation was obtained using the Pareto distribution (see Figure 1 and Table 6 columns 1 and 2). Then, using the Pareto distribution, I tested whether the parameters of the distribution are the same for low-deductible policyholders and regular-deductible policyholders. The results of a Wald test suggest that the parameters are indeed different for low-deductible policyholders and for high-deductible policyholders.

D. The Distribution of the Number of Accidents

Turning to comparing low-deductible and regular-deductible policyholders in terms of the distribution of the number of accidents, the problem again is that the data includes only accidents whose damages exceeded the level of the deductible chosen by the policyholder. The number of reported accidents is truncated; even if the record indicates zero claims for a policyholder, the policyholder might have been involved in one or more accidents whose damage was below the deductible level.

In order to estimate the distribution of the number of accidents I made the following two assumptions: (1) the distribution of the actual number of accidents, n^*, is Poisson with λ, where $\lambda=exp(z\beta)$ (z is the characteristics of the policyholder), and (2) the distribution of the number of claims, n, is Binomial with n^* and $1-q$, where q is regarded here as fixed and is equal to $P(x<d)$ (the probability that the magnitude of the claim, x, will not exceed the policyholder's deductible level, d). A value for q is obtained from the preceding subsection V.C.

Given the assumptions above, the distribution of n^*, given n, and θ (which is q, and λ) is Poisson with λq: $p(n^* \mid n,\theta) \sim Poisson(\lambda q) \quad \forall n^* >= n$ and the distribution of n is Poisson with $\lambda(1-q)$: $p(n \mid \theta) \sim Poisson(\lambda(1-q)) \quad \forall n >= 0$. Using the Newton-Raphson procedure for the number of claims, one can estimate λ_1 (which is equal to $\lambda(1-q)$) and then divide it by (1-q) to obtain an estimate for λ.

Figure 2 shows the distribution of the number of accidents and the number of claims for low-deductible policyholders and for regular-deductible policyholders, respectively. The results for the coefficients of the distribution are shown in Table 6 columns 3 and 4. The results indicate that the probability of being involved in at least one accident is 0.2 for low-deductible policyholders and only 0.146 for regular-deductible policyholders.

E. Reconsidering Prior Research: Beginning vs. Experienced Drivers

For adverse selection to arise, policyholders must have private information about their risk type that gives them an informational advantage over the insurer. Policyholders with significant driving experience have had more opportunities to gain information about their risk type.

To examine this issue, I ran the Bivariate Probit test suggested by Chiappori and Salanie (2000) separately for policyholders with less than three years of driving experience and for policyholders with three or more years of such experience. For policyholders with less than three years of driving experience, I find (see Table 7, column 1) that ρ has a positive value of 0.023 that is not statistically significant (with a standard error of 0.059). These findings are, of course, consistent with those of Chiappori and Salanie (2000) who studied policyholders with zero to two years driving experience.[9]

However, for the group of policyholders with three or more years of driving experience, a correlation between low-deductibles and more accidents does exist. For this group, ρ has a negative value of -0.06 that is statistically significant (at the 1% confidence level) and that enables rejecting the independence of the two equations (see Table 7 column 2).

These findings indicate that, although the Chiappori-Salanie results are confirmed for policyholders with 0-2 years of driving experience, they do not carry over to policies purchased by individuals with three or more years of driving experience. The evidence is thus consistent with a coverage-risks correlation, and thus with the presence of adverse selection, for such policies.

[9] Note that the results reported above are ones that do not exclude the information I have in my data on past claim history, whereas the Chiappori-Salanie findings were reached using data that did not include information on past claims history (which is why they chose to focus on young drivers). To duplicate with my data exactly what these authors did, I excluded the information that I have on the past claim history of policyholders and did the Bivariate Probit test on policyholders with less than three years of experience. Again I find that ρ, the correlation between the equation of the choice of the deductible and the equation of the occurrence of a claim, is very close to zero (0.0023) and the 95 percent interval is equal to [-0.111, 0.116]. This is similar to the Chiappori-Salanie finding that zero falls within the 95 percent confidence interval for ρ.

It is worth noting that policies sold to drivers with less than three years of experience constitute a tiny minority of the total policies sold by the companies studied by both Chiappori-Salanie and myself. There are about 1,800 such policies among the more than 200,000 policies sold by the insurer whose data I study (less than 1%). Chiappori-Salanie started with a data base of 1.1 million policies but, seeking to exclude from their testing drivers with significant past claim history for which they did not have data, concentrated on the 20,000 policies sold to policyholders with less than three years of driving experience (less than 2% of total policies).[10] Thus, because policies sold to drivers with three or more years of experience constitute the overwhelming majority of total policies, and because a coverage-accidents correlation exists for such policies, the pool of all policies is also characterized by such a correlation.

VI. LEARNING

A. Learning by Policyholders

The results already reported about how the presence of coverage-accidents correlation depends on whether the policyholder has three or more years of driving experience indicate the relevance of learning by policyholders. To take a closer look at this subject, let us consider the effects of the driving experience that customers have before joining the company. During this period, the policyholder learns information about realization of risks that (because self-reporting is not fully accurate) the insurer might not fully get when the policyholder joins.

[10] Drivers with less than three years of driving experience make up a larger percentage of total drivers. In both the French and Israeli markets, however, many young drivers do not own a car and do not purchase an automobile insurance policy; the car they use is often one for which a parent (usually with a substantial driving experience) is the main driver and the policyholder buying the insurance policy.

To test this issue, I divided all the policies into 5 sub-groups – those in which the policyholder has 0, 1, 2, 3, or 4 or more years of driving experience *not* with the insurer. And I tested, for each sub-group separately, the presence of coverage-accidents correlation.
The results, which are reported in Table 8, indicate that the correlation does not exist in the groups of customers with only 0,1, or 2 years of driving experience not with the insurer. It does exist, however, for the groups of customers who have 3 years or 4 or more years of driving experience not with the insurer. Thus, the results are consistent with the story that the longer a customer's driving experience before joining the insurer, the greater the informational asymmetry between the customer and the insurer.

Interestingly, the results that a deductible-accidents correlation does not exist for policyholders with no or little driving experience rule out the possibility that customers can obtain significant private information about automobile accident risks from introspection or from observing their performance in other dimensions of life. Only experience with one's own driving can provide significant private information about one's risk type in connection with automobile accidents.

B. Learning by Insurers: New vs. Repeat Customers

During the period in which a policyholder stays with the same insurer, both the insurer and the policyholder get to observe the realization of risks. There is, so to speak, symmetrical learning with respect to the realization of risks during this period. Assuming that the realization of risks is the only or primary source of information about risk type, as the experience of the insurer with the policyholder increases, the customer's initial informational advantage over the insurer (if any) can be expected to diminish.

To examine this issue, I test whether the association between low deductibles and higher risks is one that declines or disappears for customers with whom the insurer has had long experience. To this end, I regressed for all policyholders with a given company experience (0,1,2,3,and 4) the number of claims exceeding the regular deductible on the choice of the deductible, the exposed time of the policy, and all of the policyholder's characteristics. As Table 9 shows, the coefficient on the deductible choice decreases with the length of the company's experience with the customer.

For example, for the group of policyholders with zero years of company experience, the coefficient on the deductible choice is equal to –0.037 and is statistically significant (at the 1% confidence level). In contrast, for the group of policyholders with four years of company experience, the coefficient on the deductible choice decreases to –0.008 and is no longer statistically significant (with a standard error of 0.008). Interestingly, whereas for policyholders with company experience of 0 and 1 the coefficient on driving experience is strongly significant, it is not significant at all for policyholders with 3 or 4 years of company experience; this suggests that, once the company has three or more years of experience with the policyholder, any informational advantage of the policyholder is largely eliminated no matter how large the policyholder's driving experience.

Because the company had long experience with repeat customers only after it has operated for some time, it is worth distinguishing between learning by the insurer about individual policyholders and learning by the insurer about the general pool of policyholders it faces. Table 4 displays the results of regressions run for each one of the years in which the company operated. The results indicate that the coefficient on the deductible choice stays statistically significant (at the 1% confidence level) in each of these years. Thus, the evidence indicates that the association between low deductibles and more accidents can be eliminated only by specific learning by the

insurer about its repeat customers (and thus only with respect to such customers) but not by general learning of the insurer about the general pool of policyholders in the market.

Finally, if insurers can learn over time enough about repeat customers to eliminate the customers' initial informational advantage, does that imply that adverse selection in this market can be at most a short-run phenomenon?[11] The answer is no, because insurers in this market keep having new customers. New drivers constantly join the pool of individuals buying insurance, and, furthermore, policyholders change insurers (as will be presently documented). Therefore, since the purchasers of insurance are not all, and will not all be, repeat customers with long experience with the insurer, the association between low-deductibles and more accidents cannot be expected to vanish in the long run.

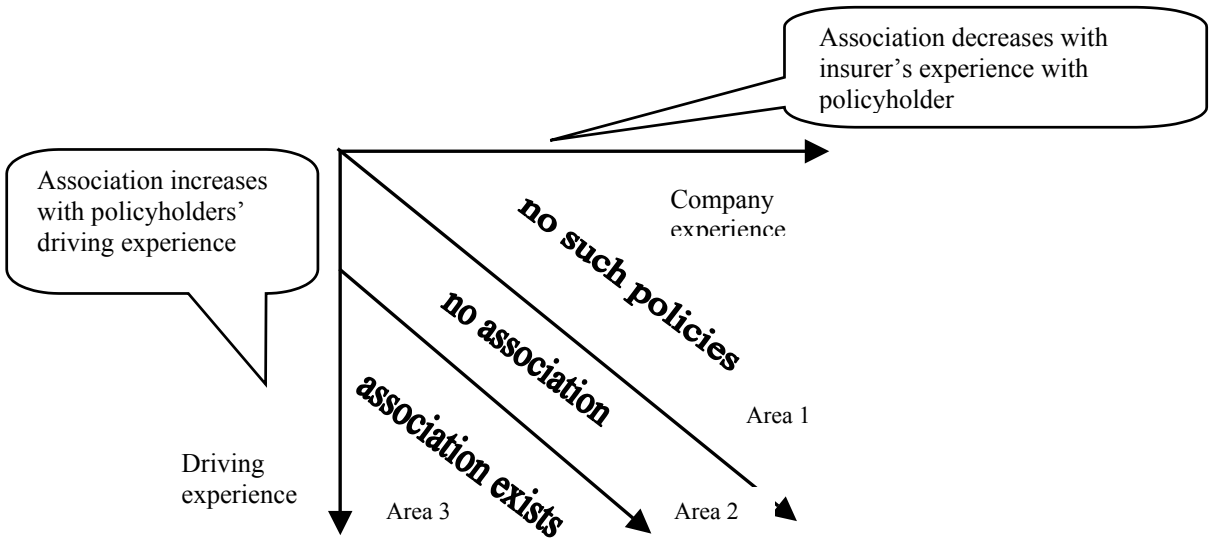
C. The Combined Effects of Policyholder and Insurer Learning

Of course, the extent to which an informational asymmetry is present in any given case depends *both* on the information (if any) learnt by the policyholder *and* on the information (if any) learnt by the insurer. The more driving experience the policyholder has before joining the insurer, other things equal, the more likely an informational asymmetry to be present. And the longer the policyholder has stayed with the insurer, other things equal, the less likely such an informational asymmetry.

Thus, if we draw two axis representing the driving experience of the policyholder and the policyholder's experience with the insurer, we can expect the presence of correlation between low-deductible and high losses from accidents to depend on these two parameters as depicted in the following diagram.

[11] For a work stressing differences between short-run and long run effects in another insurance market, see Cutler and Reber (1998).

Association Between Low Deductibles and High Risks



By definition, there are no policies in which the policyholder has been driving less years than the policyholder has been with the insurer (Area 1). When the policyholder has spent all or close to all of the policyholder's driving years with the insurer (Area 2), a coverage-accidents correlation is not expected. However, when the policyholder has a significant experience prior to joining the insurer (Area 3), a coverage-accidents correlation is not expected.

To test this prediction, I divided all policies in the data into 25 sub-groups, with each sub-group including all the policies with a given number of years of driving experience by the policyholder and a given number of years of experience with the policyholder by the insurer. I then tested, for each group separately, the presence of correlation between deductible choice and risk type.

The results of these regressions, which are reported in Table 10, are largely consistent with the tested prediction. To illustrate, the column with sub-groups of new customers with zero years

of experience with the insurer indicates that a correlation appears for such customers once they have three or more years of driving experience; in contrast, the column with sub-groups of customers that have four years of experience with the insurer indicates that a correlation appears for such customers only when they have six or more years of driving experience (and thus two or more years of driving experience not with the insurer). The row with sub-groups of policyholders that have three years of driving experience indicates that the correlation disappears for customers who have spent all these three years with the insurer; in contrast, the row with sub-groups of policyholders with five or more years of driving experience indicates that the correlation disappears only for customers who have spent at least four years with the insurer.

D. Learning and Differences in Information among Insurers

I now turn to test whether the information obtained by insurers about their repeat customers produces a difference in information among insurers with respect to these customers. To start with, I wish to observe that some of the results obtained in the preceding section are already ones that are predicted by this story. We have seen that the effect of driving experience on the presence of a coverage-risks correlation depends on whether this driving experience was acquired while the policyholder was insured by the insurer or by other insurers. Driving experience that has been spent with the insurer reduces the correlation between low-deductibles and accidents, but driving experience with other insurers does not (on the contrary). This is not what would be expected if the information of other insurers that have served new customers in the past were fully available to the insurer.

1. Inside vs. Outside Record as Predictors of Subsequent Performance

The story that we consider is one in which an insurer can fully count on its own records of the past claim history of repeat customers but cannot fully count on customers' reports about

their past claim history with other insurers. On this story, the actual record of new customers is systematically worse than the reported record. This implies that the number of claims for new customers would be higher than what would be predicted assuming that the customers' self-reporting of past claim history were accurate.

To test this, I took (i) all the new customers, and (ii) all the customers who have been for three years with the insurer. The number of claims in the past three years which is included in the insurer's data is based on self-reporting for the first group and on the insurer's own records for the second group. I regressed the total costs to the insurer from each policy on all the characteristics of the policyholder including the number of past claims and on whether the customer is a new customer or a repeat customer. The results, which are displayed in Table 11A, indicate that the number of claims is higher, in a statistically significant way (at the 1% confidence level) for the new customers.[12] In unreported regressions, I find that this result holds when I run separate regressions for customers with one, two, or three or more claims in their past.

2. Departing Customers

Another implication of the considered model is that departing customers will be disproportionately ones with a poor past record. Customers with such a record could benefit more from switching to a new insurer who would have less information about their past history.

To take a first look at this issue, I calculated some summary statistics which are displayed in Table 11B. They show that, in each year and for each cohort and deductible group, the customers that departed had more claims in the preceding year than the staying customers.

[12] In particular, a new customer who reports a clean record of no claims in the past three years will be charged a higher regular premium than a repeat customer who has been with the company for three years and had such a clean record in those three years so long as the prior record reported by the repeat customer when joining the company three years ago was not an especially bad one with two or more claims.

To test this issue, I created a dummy variable that was equal to 1 if the customer decided to stay in the company for another year and 0 otherwise. The decision whether to stay was regressed on whether the customer had claims in the year prior to this decision, and on all the characteristics of the customer including the deductible choice in the preceding year. The coefficient on whether the customer had at least one claim in the year prior to the decision whether to stay was negative and statistically significant.

The results, which are displayed in Table 11C, indicate that the probability of staying with the insurer for another year is smaller by 0.1 (at the 1% confidence level) for customers who had at least one claim in the preceding year than for customers who had no claims in that year. For example, whereas customers in any given year have on average a probability of 0.7 of staying with the insurer for another year, customers who had a claim in the preceding year had a probability of only 0.6 of staying with the insurer for another year.

These results suggest that departing customers tend to be disproportionately ones whose record with the company included claims. One might still raise the question of whether departing customers are in fact high-risk types or individuals who had more claims simply due to bad luck. To examine this question, I looked at the realization of risks for customers who stayed and tested whether those who stayed had subsequently good performance compared with the general pool of customers.

I ran, for the fifth year of the company's operations, two regressions. For each choice of deductible (regular/low), I regressed the number of claims on all the observable characteristics including the number of years of experience that the insurer has had with the customer (company experience). The results, which are reported in Table 11D, show that the coefficient on the company experience is negative and statistically significant. The longer the company's experience with the policyholder, the lower the likelihood that the policyholder will have claims.

The regression shows that, for low-deductible policyholders, each year of experience reduces the number of claims by 0.013 (at the 1% confidence level). This amounts to 6% of the number of claims that low-deductible policyholders have. Repeat customers, then, are associated with lower losses.[13]

3. Profits on New and Repeat Customers

I now turn to the last testable implication discussed earlier of models in which an insurer obtains over time information about its repeat customers that other insurers do not fully have. In these models, the insurer will have some market power with respect to repeat customers that the insurer will identify as low-risk types. Whereas the insurer knows them to be of a low-risk type, another insurer to which such customers might turn will not know this for sure. Therefore, the insurer will be able to over-charge repeat customers with excellent past records, and, anticipating this possibility, the insure will be willing to under-charge new customers. The testable implication, then, is that the insurer will make higher profits – or, equivalently, will have lower loss ratios – on repeat customers than on new customers.

To test this prediction, I estimated an expected loss ratio for each policyholder. I first generated a measure of expected loss ratio, LRETAC, which is equal to the expected total annual costs divided by the yearly premium.[14] I regressed each of the expected loss ratios on the

[13] This effect seems to be larger for individuals who chose low deductible than for those who chose regular deductible. For example, each year of experience decreases the number of claims by 0.013 (0.003) for individuals who chose low deductible and by 0.003 (0.0008) for individuals who chose regular deductible. And each year of experience decreased the total insurance payments by the insurer by 152 NIS for individuals who chose low deductible and by 35 NIS for individuals who chose regular deductible.

[14] The total annual costs used in calculating the LRETAC include in addition to the total insurance payments (if any) made to the policyholder also incurred administrative costs in the event of an accident and reimbursements of premia paid in the event of departure prior to year-end. (See Appendix I for a precise definition.) I checked and found that the results hold for alternative standard measures of loss ratio.

insurer's experience with the customer, the insurer's s year of operation, and the choice of the deductible.

The results, which are reported in Table 12, indicate that, the expected loss ratio deceases, in a statistically significant way (at the 1% confidence level), with the insurer's experience with the customer. For example, an increase of one year in the insurer's years of experience with the customer reduces the expected loss ratio by 1 percentage point (at the 1% confidence level).[15]

E. Remarks on Moral Hazard

As noted in the introduction, finding a correlation between deductibles and risk type is also a necessary condition for the presence of moral hazard, and my findings that such a correlation exists indicate that the rejection of moral hazard in this market by recent work was premature. Thus, it might be suggested that the findings of this paper, while consistent with adverse selection, could be all produced by moral hazard and that the considered market thus might still not involve adverse selection.

It is therefore worth noting that, although the presence of the coverage-accidents correlation identified in Section V is consistent with moral hazard, some of the findings in this Section VI do not seem readily explainable by a simple moral hazard story. For example, if the coverage-risks correlation results from low deductibles leading to lower precautions, why would such an effect not exist also for young drivers? Are incentives not important with respect to such drivers? Perhaps some driving experience is needed for policyholders to know what precautions to take. But if this were the case, why would no correlation exist with respect to policyholders that have been with the insurer for a long period of time?

[15] It is worth noting that higher profits on new and repeat customers are also consistent with a model with no adverse selection but rather switching costs that provide insurers with market power over customers that they already have. Note that a model with switching costs and no adverse selection, however, cannot readily explain another finding of this section – that policyholders switching to other insurers are disproportionately ones with a bad past record.

Furthermore, if there is no informational asymmetry between insurers and new customers, why do policyholders who leave their insurer tend to be ones with a poor past record? If insurers all know what customers know about their own risk type, why do such policyholders have systematically more to gain from switching insurers? Also, if insurers do not obtain market power with respect to some repeat customers by having an informational advantage with respect to them over other insurers, why are insurers willing to make lower profits on new customers who are not committed in any way to stay with them? All these are questions that are not readily explainable by a simple moral hazard story.

In any event, a full testing for moral hazard is beyond the scope of this paper and is pursued separately (Cohen and Deheija (2001)). This paper seeks to disentangle moral hazard and adverse selection using the data considered here and a testing approach suggested by recently by Chiappori, Heckman, and Pinquet (2001) for testing for moral hazard. These authors suggest basing such testing on the different predictions that adverse selection and moral hazard stories have for the dynamics of accident probability in the presence of "experience rating" – specifically, for whether having an accident in a given period increases or decreases the probability of the policyholder's having an accident in subsequent years. Using this test, we reject the moral hazard story and thus conclude that the coverage-accidents correlation cannot be fully explained by moral hazard rather than adverse selection.

VII. Conclusion

Using a unique and rich database, which includes all the data that an insurer had about its customers, this paper has tested the predictions of adverse selection models. Consistent with the presence of adverse selection, I found that an insurance menu with different deductibles leads to separation between high- and low-risk individuals. Low deductibles are associated with more

accidents, higher losses from accidents, and a different distribution of the number of accidents and the damage in the event of an accident.

Whether any informational asymmetry is present (an, if so, what its magnitude is) might change over time, as parties obtain more information. Consistent with the presence of such learning, I have found that the correlation between low deductibles and more accidents exists only for policyholders drivers that are sufficiently experienced with their own driving, and that it diminishes and eventually disappears when a policyholders stays with an the insurer for a long period of time.

Finally, the evidence is also consistent with insurers' obtaining information about their repeat customers that other insurers to which such customers might turn would not fully have. In particular, the costs to the insurer of new customers are higher than what would be predicted if the reports by these customers regarding their past claim history were as reliable as the insurer's records about the past claim history of repeat customers. Furthermore, consistent with insurers knowing more about their customers than other insurers, customers who leave their insurer are disproportionately ones with a poor past record, and insurers make higher profits on their repeat customers than on new customers.

REFERENCES

All-Industry Research Advisory Council (AIRAC), *Evaluation of Motor Vehicle Records as a Source of Information on Driver Accidents and Convictions*, Oak Brook (1984), III.: AIRAC

Akerlof, George. A. (1970), "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism, " *Quarterly Journal of Economics*, Vol. 84, pp. 488-500.

Arrow, Kenneth J. (1963), "Uncertainty and the Welfare Economics of Medical Care, " *American Economic Review*, Vol. 53, pp. 941-973.

Chiappori, Pierre-Andre (2000), "Econometric Models of Insurance under Asymmetric Information," in *Handbook of Insurance* (G. Dionne, ed.), 365-393.

Chiappori, Pierre-Andre, James Heckman, and Jean Pinquet (2001), "Testing for Moral Hazard on Dynamic Insurance Data," Working Paper, University of Chicago and Universite Paris X – Nanterre..

Chiappori, Pierre-Andre, and Bernard Salanie (1997), "Empirical Contract Theory: The case of Insurance Data," *European Economic Review*, Vol. 41, pp. 943-950.

Chiappori, Pierre-Andre, and Bernard Salanie (2000), "Testing for Asymmetric Information in Insurance Markets, " *Journal of Political Economy*, Vol.108, No. 1, pp. 56-78.

Cohen, Alma and Rajeev Dehejia (2001), "Testing for Moral Hazard in the Automobile Insurance Market," work in progress, Harvard University and Columbia University.

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

Crocker, Keith J. and Arthur Snow (1985), "The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information, " *Journal of Public Economics*, Vol.26, pp. 207-219.

Cutler, David M. and Sarah Reber (1998), "Paying for Health Insurance: The Trade-off Between Competition and Adverse Selection, " *The Quarterly Journal of Economics*, pp. 433-466.

Cutler, David M. and Richard J Zeckhauser (2000), "The Anatomy of Health Insurance, " forthcoming in A. Cutler and J. Newhouse, eds., *Handbook of Health Economics* .

Cummins, J. D., and J. Van Derhei, (1979), "A Note on the Relative Efficiency of Property-Liability Insurance Distribution System, " *Bell Journal of Economics*, Vol. 10, pp. 709-719.

D'arcy, Stephen, P., and Neil, A. Doherty (1990), "Adverse Selection, Private Information, and Lowballing in Insurance Markets, " *Journal of Business*, Vol.63, No. 2, pp. 145-164.

Dahlby, Bevan G. (1983), "Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance." *Journal of Public Economic* Vol.20, pp. 207-219.

Dahlby, Bevan G. (1992), "Testing for Asymmetric Information in Canadian Automobile Insurance." In *Contributions to Insurance Economics* (Georges Dionne ed.).

De Meza, David, and David C. Webb (2001), "Advantageous Selection in Insurance Markets, " *Rand Journal of Economics*, Vol. 32, No. 2, pp. 249-262.

Dionne, Georges (1983), "Adverse Selection and Repeated Insurance Contracts, " *Geneva Papers on Risk and Insurance*, Vol. 8, pp. 316-332.

Dionne, Georges, and Neil A. Doherty (1994), "Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from Insurance Market," *Journal of Political Economy*, Vol. 102, No. 2, pp.210-235.

Dionne, Georges, Doherty, Neil and Nathalie Fombaron (2000), "Adverse Selection in Insurance Markets," in *Handbook of Insurance* (G. Dionne, ed.), pp. 185-244.

Dionne, Georges, Gouriéroux Christian, and Charles Vanasse (2001), "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment," *Journal of Political Economy*, Vol. 109, No. 2, pp. 444-453.

Dionne, Georges and Philippe Lasserre (1985), "Adverse Selection, Repeated Insurance Contracts and Announcement Strategy," *Review of Economics Studies*, Vol. 52, pp. 719-723.

Dionne Georges and Philippe Lasserre (1987)," Adverse Selection and Finite-Horizon insurance contracts, " *European Economic Review*, Vol. 31, pp. 843-861.

Dionne, Georges, Gouriéroux Christian, and Charles Vanasse (1992), "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information," *Journal of Applied Econometrics* 7, 149-165.

Finkelstein, Amy, and James Poterba (2000), "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market, " *National Bureau of Economic Research*, Working Paper 8045.

Fombaron, Nathalie (1997), "No-commitment and Dynamic Contracts in Competitive Markets with Asymmetric Information, Working Paper, Thema.

Insurance Research Council (1991), *Adequacy of Motor Vehicle Records in Evaluating Driver Performance* (Insurance Research Council, Oak Brook, Illinois).

Nilssen, T. (2000), "Consumer Lock-in with Asymmetric Information, *International Journal of Industrial Organization*, Vol. 18, pp 641-666.

Friedman, M. Benjamin and Mark J. Warshawaski, (1990), "The Cost of an Annuities: Implication for Saving Behavior and Bequests, " *Quarterly Journal of Economics*, Vol. 105, pp. 135-54.

Genesove, David (1993), "Adverse Selection in the Wholesale Used Car Market, " *Journal of Political Economy*, Vol. 101, No. 4, pp. 644-665.

Grossman, Herschel I. (1979), "Adverse Selection, Dissembling, and Competitive Equilibrium," *The Bell Journal of Economics*, Vol. 10, pp. 336-343.

Hosios. J. Arthur, and Michael Peters (1989), "Repeat Insurance Contracts with Adverse Selection and Limited Commitment," *Quarterly Journal of Economics*, pp. 289-253.

Kunreuther, Howard and Mark V. Pauly (1985), "Markets equilibrium with Private Knowledge: An Example," *Journal of Public Economics*, Vol. 26, pp. 269-288.

Miyazaki Hajime (1977), "The Rat Race and Internal Labor Markets," *The Bell Journal of Economics* Vol. 8, No. 2, pp. 394-418.

Puelz, Robert, and Arthur Snow (1994), "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market," *Journal of Political Economy*, pp. 236-257.

Raviv, Arthur (1979), "The Design of an Optimal Insurance policy," *American Economic Review*, pp. 84-96.

Riley, John G. (1979), "Informational Equilibrium," *Econometrica*, Vol. 47, No. 2, pp. 331-360.

Riley, John C. (2001), "Silver Signals: Twenty-Five Years of Screening and Signaling," *Journal of Economic Literature*, Vol. 39, 432-478.

Rothschild, Michael and Joseph E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly journal of Economics* Vol. 90, pp. 629-649.

Shavell, Steve (1979), "On Moral Hazard and Insurance," *Quarterly Journal of Economics*, pp. 541-562.

Spence, Michael A. and Richard J. Zeckhauser (1971), "Insurance, Information, and Individual Action," *American Economic review*, Vol. 61, pp. 380-387.

Spence, Michael A. (1978), "Product Differentiation and Performance in Insurance Markets," *Journal of Public Economics*, Vol. 10, pp. 427-447.

Stiglitz, Joseph E. (1977), "Monopoly, Nonlinear pricing, and Imperfect Information; the insurance Market, " *Review of Economics Studies*, Vol. 44, pp. 407-430.

The Royal Swedish Academy of Sciences, Markets with Asymmetric Information (advanced information), available at www.nobel.se/economics/laureates/2001/ecoadv.pdf.

Wilson, Charles A. (1977), "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, Vol. 16, pp. 167-207.

Winter, Ralph (2000), "Optimal Insurance under Moral Hazard," in *Handbook of Insurance* (G. Dionne, ed.), 155-184.

APPENDIX I

List of Variables (in Alphabetical Order)

Academic education	Equal to 1 if the policyholder has a university degree and equal to 0 otherwise
Actual profit	Earned premium minus payments paid by the insurer plus expenses on processing claims
Age	Age of policyholder
Age-gender	Interaction between age and sex
Car model year	Car model year
CC	Size of engine
Claim Index	Equal to 1 if the number of claims in the current year is greater than 1 and equal to 0 otherwise
Claim1 - Claim4	Number of claims occurs in the first (second/third and forth) year that the policyholder was enrolled in the insurance company
Company experience	Equal to the number of years that the policyholder has been with the insurer
Calendar year	Equal to the calendar year
Damage	The amount of the damages reported in a claim
Deductible level	Equal to 1 if the level of the deductible is regular and equal to 0 otherwise
Driving experience	Length (in years) that the policyholder has had a driving license
Earned premium	[Premium] * [Period]
Immigrant index	Equal to 1 if the policyholder is an immigrant
Immigration year	The year in which an immigrant policyholder immigrated to Israel
Expected profit	Equal to yearly premium minus the expected Total Actual Payment
Gender	The policyholder's gender
LRETIP	Expected [Total insurance payments] divided by [Earned premium]
LRETAC	Expected [Total Annual Cost] divided by [Premium]
Main car	Equal to 1 if the car is used as main car and equal to 2 otherwise
No experience last year	Equal to 1 if the policyholder did not have a driver license in the year prior to joining the insurer
No experience three years ago	Equal to 1 if the policyholder did not have a driver license during the year that was three years prior to joining the insurer
No experience two years ago	Equal to 1 if the insured did not have a driver license in the year taking place two years prior to joining the insurer
Number of claims	Number of claims submitted during the life of the policy
Number of claims last year	Number of claims occurring in the year prior to joining the insurer
Number of claims three years ago	Number of claims occurring three years prior to joining the insurer
Number of claims two years ago	Number of claims two year before the policyholder joined the insurer
Number of drivers	Equal to 1 if only the policyholder and the policyholder's spouse use the car and 0 otherwise
Period	Length of the period covered by the policy (in years)
Premium	Yearly premium

Regular deductible	Equal to 1 if the policyholder chose a regular deductible and 0 otherwise
Regular Premium	The premium the policyholder were quoted for a regular deductible (and was charged if the policyholder chose such a deductible)
Single	Equal to 1 if the policyholder is single and to 0 otherwise
Social and private use	Equal to 1 if the car is used for social and private needs and equal to 0 otherwise
Stayed	Equal to 1 if the policyholder was insured by the company in the preceding year
Total annual costs	All payments resulting from the issuance of the policy = payments made to the policyholder in connection with claims + administrative expenses in handling a claim + reimbursements of premia payments made in the event of early departure
Total insurance payments	Payments made to the policyholder in connection with claims
Value of the car	The value of the car
Young driver age	The age of the youngest driver who is allowed to use the car
Young driver experience	The driving experience of the youngest driver using the car
Young driver gender	The youngest driver's gender
Young driver index	Equal to 1 if the a young driver used the car and equal to 0 otherwise.
Young last year	Equal to 1 if the policyholder was considered a young driver in the year prior to joining the insurance company
Young three years ago	Equal to 1 if the policyholder was considered a young driver three years prior to joining the insurance company
Young two years ago	Equal to 1 if the policyholder was considered a young driver two years prior to joining the insurance company

APPENDIX II

1) Claim frequency – The total number of claims made by customers divided by the number of policies weighted by the exposed time of the policy, which is the time (in year unit) that the policy was in effect.

$$\frac{\sum_i number\ \ of\ \ claims_i}{\sum_i \exp osetime_i}$$

2) Loss ratio (damage) – The sum of all damages incurred to policyholders divided by the sum of the insurer's earned premium, which is the sum of all annual premium weighted by the exposed time of the policy.

$$\frac{\sum_i damage_i}{\sum_i premium_i * \exp osetime_i}$$

3) Loss ratio (cost) – The sum of all the payments made to customers by the insurer divided by the insurer's total earned premium.

$$\frac{\sum_i payment_i}{\sum_i premium_i * \exp osetime_i}$$

4) Average premium – The sum of all earned premium divided by the exposed time

$$\frac{\sum_i premium_i * \exp osetime_i}{\sum_i \exp osetime_i}$$

5) Average damage per policy – The sum of the damages incurred to the policyholders divided by the exposed time.

$$\frac{\sum_i damage_i}{\sum_i premium_i * \exp osetime_i}$$

6) Average cost per policy – The sum of the payments made by the insurer divided by the total exposed time of policies.

$$\frac{\sum_i payment_i}{\sum_i premium_i * \exp osetime_i}$$

7) Average damage per claim – The sum of the damages incurred to policyholders divided by the number of claims.

$$\frac{\sum_i damage_i}{\sum_i number\ \ of\ \ claims_i}$$

8) Average cost per claim – The sum of all the payments made by the insurer divided by the number of claims.

$$\frac{\sum_i payment_i}{\sum_i number\ \ of\ \ claims_i}$$

Table 1: Descriptive Statistics

Type of policies	All the policies		Regular deductible policies		Low deductible policies	
Variable	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
Buyer Demographics characteristics:						
Age	42.5	12.54	42.6	12.47	42.4	12.78
Gender	1.32	0.47	1.31	0.47	1.32	0.46
Single	0.13	0.33	0.13	0.33	0.13	0.33
Academic education	0.26	0.04	0.25	0.43	0.32	0.46
Buyer's car characteristics:						
CC	1,565	3750	1,566	380	1,559	358
Car model year	1992.9	3.20	1993.02	3.20	1992.6	3.20
Value of the car	61,932	34,998	60,997	34,780	65,202	35,559
Social and private use	1.08	0.27	1.08	0.27	1.07	0.26
Main car	1.15	0.36	1.15	0.36	1.14	0.35
Buyer 's driving characteristics:						
Driving experience	19.03	10.13	18.97	10.12	19.2	10.13
Number of claims last year	0.07	0.27	0.07	0.26	0.08	0.28
Number of claims two years ago	0.05	0.22	0.04	0.22	0.05	0.23
Number of claims three years ago	0.04	0.20	0.04	0.19	0.04	0.19
Claim1	0.052	0.023	0.05	0.22	0.07	0.28
Claim2	0.026	0.17	0.02	0.15	0.04	0.21
Claim3	0.01	0.11	0.008	0.10	0.02	0.15
Claim4	0.003	0.06	0.002	0.05	0.006	0.08
Number of claims	0.19	0.45	0.18	0.44	0.23	0.50
Damage	2,136	8,718	2,032	8,561	2,501	9,239
Total insurance payments	1,538	7,301	1,455	7,163	1,824	7,755
Average cost	1,383	6,725	1,618	7,082	1,316	6,618
Total actual payment	2,028	7,526	1,966	7,382	2,246	8,004
Number of drivers	0.25	0.43	0.25	0.44	0.22	0.41
Young driver index	0.17	0.04	0.17	0.38	0.18	0.38
Stayed	0.7	0.46	0.7	0.46	0.69	0.46
N	213,660		166,118		47,542	

Table 2: Premium and Total Annual Costs as a Function of Characteristics

Dependent Variable:	Premium		Total annual cost	
OLS	Coef.	std.	coef.	std.
Buyer demographic characteristics:				
Age	1.38***	0.34	-1.19	5.30
Gender	92.74***	9.40	-37.58	144.62
Agesex	-3.72***	0.22	-2.35	3.37
Single	98.03***	3.84	76.64	59.02
Academic education	-55.17***	2.78	-218.60***	42.66
Buyer's car characteristics:				
CC	0.21***	0.004	-0.36***	0.06
Car model year	-13.11***	0.58	-78.39***	8.85
Value of the car	0.02***	0.00006	-0.021***	0.001
Main car	-54.60***	3.36	-86.02*	51.68
Buyer's driving experience:				
Driving experience	-4.75***	0.23	-19.79***	3.47
Number of claims last year	447.91***	4.45	757.46***	68.42
Number of claims two years ago	303.95***	5.40	633.76***	82.93
Number of claims three years ago	254.77***	5.96	500.47***	91.48
No experience last year	196.35***	14.41	186.13	221.41
No experience two years age	189.57***	16.22	419.24*	249.15
No experience three years age	328.74***	10.26	120.49	157.57
Number of drivers	-88.20***	2.80	-152.21***	43.00
Young driver age	-283.60***	3.50	-222.14***	53.80
Young driver experience	-263.28***	4.22	142.44***	64.90
Young driver gender	-172.83***	5.75	-339.12***	88.37
Young driver index	2194.64***	12.75	2040.01***	195.86
Young driver last year	235.87***	27.65	785.71	424.82
Young driver two years age	223.34***	30.81	529.52	474.15
Young driver three years ago	274.34***	23.40	-382.95	359.45
Claim1	443.19***	5.41	564.48***	83.15
Claim2	346.30***	7.850	432.25***	120.66
Claim3	245.29***	12.50	183.00***	191.98
Claim4	250.79***	25.42	666.74*	390.47
Other:				
Company experience	-74.26***	1.00	-62.67	15.43
Time Fixed Effect	YES		YES	
N	166,116		166,116	
Adj-R^2	0.71		0.79	

***,**,* - Significant at 1%, 5%, and 10% confidence level, respectively

Table 3: Summary Statistics

Level of Deductible	"Low"	"Regular"	"High"	"Very high"
Percentage of choosing	21.96%	76.72%	0.74%	0.6%
Claim frequency	27.76%	21.33%	14.05%	11.37%
Loss Ratio (damage)	99.59%	93.04%	91.7%	63.62%
Loss Ratio (cost)	72.62%	66.65%	65.29%	42.34%
Average premium	2,853	2,623	2,085	1,920
Average Damage per policy	2,841	2,440	1,911	1,222
Average cost per policy	2,071	1,748	1,361	813
Average damage per claim	10,233	11,443	13,600	10,750
Average cost per claim	7,462	8,198	9,683	7,154

- See Appendix II for the exact definition of each of the above terms.

Table 4: The Association between Deductible Choice and Accidents

Column		**1**	**2**
Company year:	**Dependent variable: OLS**	**Number of claims**	**Total insurance payments**
	Deductible level	-0.04***	-241.44**
		(0.007)	(126.6)
1			
	N	21,715	21,715
	Adjusted-R^2	0.0073	0.032
	Deductible level	-0.03***	-286.14***
		(0.007)	(92.1)
2			
	N	38,201	38,201
	Adjusted-R^2	0.0076	0.0347
	Deductible level	-0.03***	-236.89***
		(0.005)	(83.3)
3			
	N	45,760	45,760
	Adjusted-R^2	0.0091	0.034
	Deductible level	-0.03***	-315.19***
		(0.005)	(80.5)
4			
	N	50,571	50,571
	Adjusted-R^2	0.0073	.038
	Deductible level	-0.03***	-229.28***
		(0.005)	(63.8)
5			
	N	57,410	57,410
	Adjusted-R^2	0.0153	0.008

- I only report the coefficients of interest
- ***, **, * - Significant at 1%, 5%, and 10% confidence level, respectively

Table 5: Bivariate Probit for the Choice of Deductible and the Occurrence of a Claim

Dependent variable:	Regular Deductible		Claim Index	
	coef.	std.	coef.	std.
Buyer's demographic characteristics:				
Age	5.34e-05	4.8e-04	0.001**	0.0005
Gender	-0.049***	0.07	-0.007	0.008
Single	-0.022**	0.010	0.024**	0.011
Academic education	-0.050***	0.007	-0.019**	0.008
Buyer's car characteristics:				
CC	-2.08e-06	1.12e-05	3.75e-05***	1.1e-05
Car model year	-0.0165***	0.002	-0.007***	0.0016
Value of car	-3.8e-07**	1.6e-07	1.14e-06***	1.6e-07
Private social use	0.095***	0.012	0.046***	0.012
Main car	0.047***	0.009	-0.065***	0.010
Buyer's driving characteristics:				
Driving experience	0.002***	0.001	-0.004***	0.001
Number of claim last year	-0.099***	0.012	0.159***	0.012
Number of claims two years ago	-0.089***	0.014	0.113	0.014
Number of claims three years ago	0.0117	0.016	0.140***	0.016
No experience last year	0.168***	0.045	0.041	0.041
No experience two years ago	-0.055	0.050	0.049	0.047
No experience three years ago	0.204***	0.032	0.017	0.029
Young driver last year	-0.073	0.063	-0.023	0.064
Young driver two years ago	0.028	0.070	0.021	0.072
Young driver three years ago	0.0432	0.053	0.040	0.055
Claim1	-0.239***	0.013	0.158***	0.014
Claim2	-0.276***	0.017	0.157***	0.020
Claim3	-0.275***	0.026	0.085***	0.032
Claim4	-0.320***	0.049	0.245***	0.062
Young driver index	0.239***	0.034	0.512***	0.033
Young driver age	-0.024**	0.009	-0.039***	0.009
Young driver experience	-0.054***	0.011	-0.070***	0.011
Young driver gender	-0.060***	0.015	-0.070***	0.015
Number of drivers	0.135***	0.008	-0.029***	0.008
Others:				
Company experience	-0.326***	0.004	-0.019***	0.003
Period	-0.115***	0.014	0.220***	0.014
Time Fixed Effect	YES		YES	
ρ	-0.057***	0.0047		
	[95% conf. interval]			
	[-0.067, -0.048]			
N	213657		213,657	

- standard errors adjusted for clustering on policy base
- ***, **, * - Significant at 1%, 5%, and 10% confidence level, respectively

Table 6: Simulating the Distributions of the Damages in the Event of an Accident and of the Number of Accidents

Column	1		2		3		4	
Dependent variable:	**Damage**		**Damage**		**Number of accidents**		**Number of accidents**	
Deductible level:	**Low**		**Reg**		**Low**		**Reg**	
	Coef.	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**
Buyer's demographic characteristics:								
Age	-35.135	-0.658	-35.301	-0.891	0.001	0.309	0.002	1.071
Gender	-495.713	-0.356	-431.881	-0.402	-0.062	-0.824	0.002	0.048
Age-gender	-13.000	-0.406	-17.609	-0.711	0.001	0.782	-0.0005	-0.407
Single	-329.860	-0.547	-663.090	-1.511	0.051	1.694	0.037	1.888
Academic education	-851.618	-2.064	-883.480	-2.805	-0.063	-2.968	-0.019	-1.347
Emigrant year	6.562	1.948	-2.702	-0.676	-0.0002	-0.761	-0.001	-3.238
Emigrant dummy	12562	1.895	-5411.100	-0.688	-0.475	-0.747	-1.072	-3.353
Buyer's car characteristics:								
CC	1.532	2.099	1.645	3.329	0.0001	3.109	0.0001	5.599
Car model year	-1.466	-0.281	11.233	2.528	-0.001	-2.671	-0.001	-5.098
Value of car	0.042	5.685	0.078	13.985	7.12E-07	2.210	9.76E-07	4.883
Main car	-167.768	-0.226	-2266.200	-4.990	0.068	1.852	0.055	2.483
Buyer's driving experience:								
Driving experience	-8.619	-0.243	-53.635	-2.163	-0.007	-3.613	-0.008	-6.586
Number of claims last year	1104.500	1.748	1048.000	2.126	0.240	7.932	0.247	12.148
Number of claims two years ago	876.900	1.078	1147.000	1.900	0.138	3.539	0.203	7.951
Number of claims three years ago	1129.200	1.206	543.850	0.856	0.211	4.636	0.230	8.309
No experience last year	-801.100	-0.429	-48.975	-0.041	0.011	0.122	0.139	2.763
No experience two years ago	-380.600	-0.169	896.411	0.633	-0.031	-0.308	0.003	0.052
No experience three years ago	581.200	0.342	-438.633	-0.418	0.138	1.827	0.050	1.096
Young driver age	-201.533	-0.965	-290.248	-1.896	-0.025	-2.357	-0.024	-3.636
Young driver experience	949.527	2.595	498.871	1.807	-0.065	-3.370	-0.096	-8.050
Young driver gender	-1721.800	-2.033	-273.158	-0.424	-0.077	-1.858	-0.127	-4.760
Young driver index	4975.500	1.235	6612.100	2.194	1.005	4.908	1.218	9.546
Claim1	752.627	1.101	133.133	0.226	0.235	7.239	0.255	9.904
Claim2	-117.113	-0.135	-631.729	-0.738	0.270	6.083	0.236	5.866
Claim3	-462.752	-0.327	-1583.200	-1.045	0.112	1.417	0.146	1.956
Number of drivers	371.391	0.763	-421.538	-1.284	0.01	0.412	-0.065	-4.280
Others:								
Company experience	381.503	1.444	476.849	2.719	-0.089	-6.621	-0.058	-7.333
Period	--------	--------	--------	--------	0.299	6.511	0.327	11.990
Time Fixed Effect	YES		YES		YES		YES	

Table 7: Bivariate Probit for Young and Experienced Drivers

Column:	1	2	3	4
Dependent variable: Driving experience:	**Regular Deductible 0-2 years**	**Claim Index**	**Regular Deductible 3 or more years**	**Claim Index**
ρ	0.023 (0.059)	[95% conf. interval] [-0.09,0.14]	-0.058*** (0.004)	[95% conf. interval] [-0.07,-0.05]
Time Fixed Effect	YES		YES	
N	1,774		211,866	

- I only report the coefficient of interest
- standard errors adjusted for clustering on policy base
- ***,**,* - Significant at 1%, 5%, and 10% confidence level, respectively

Table 8: The Effect of Driving Experience before Joining the Insurer

Dependent variable: Number of claims

Number years of experience not in the insurance company	0	1	2	3	4
Deductible level	0.04 (0.07)	-0.026 (0.053)	-0.045 (0.04)	-0.071*** (0.025)	-0.05*** (0.002)
Time fixed effect	YES	YES	YES	YES	YES
N	393	778	1421	2318	208363
Adjusted R^2	0.1073	0.0833	0.0446	0.0450	0.0236

- I only report the coefficient of interest
- ***,**,* - Significant at 1%, 5%, and 10% confidence level, respectively

Table 9: The Effect of Company Experience

Dependent variable: Number of Claims

Company years of experience	0	1	2	3	4
Deductible level	-0.037*** (0.004)	-0.036*** (0.005)	-0.034*** (0.006)	-0.021*** (0.006)	-0.008 (0.008)
log(driving experience)	-0.0377*** (0.005)	-0.017*** (0.006)	-0.014* (0.008)	-0.011 (0.010)	-0.01 (0.010)
Period	0.093*** (0.006)	0.011 (0.009)	0.009 (0.010)	0.029** (0.012)	0.092*** (0.013)
Time Fixed Effect	YES	YES	YES	YES	YES
N	97,976	56,628	31,423	15,634	8,558
Adj-R^2	0.0218	0.0217	0.0218	0.0230	0.026

- I only report the coefficients of interest
- ***,**,* - Significant at 1%, 5%, and 10% confidence level, respectively

Table 10: The Two Dimensions of Experience

Years of driving experience \ Years of company experience	0	1	2	3	4
0	NO Number of Observation: 200				
1	NO Number of Observation: 344	NO Number of Observation: 147			
2	NO Number of Observation: 570	NO Number of Observation: 158	NO Number of Observation: 117		
3	-0.126 (0.04) Number of Observation: 947	-0.103 (0.07) Number of Observation: 286	NO Number of Observation: 95	NO Number of Observation: 93	
4	-0.104 (0.04) Number of Observation: 1398	-0.06 (0.04) Number of Observation: 505	NO Number of Observation: 40	-0.28 (0.08) Number of Observation: 72	NO Number of Observation: 77
5	-0.104 (0.03) Number of Observation: 2052	-0.20 (0.042) Number of Observation: 768	-0.072 (0.06) Number of Observation: 336	-0.095 (0.08) Number of Observation: 135	NO Number of Observation: 109
6+	-0.06 (0.004) Number of Observation: 92523	-0.054 (0.005) Number of Observation: 54796	-0.057 (0.005) Number of Observation: 30695	-0.036 (0.006) Number of Observation: 15369	-0.015 (0.007) Number of Observation: 8218

All the reported figures are statistically significant in the 1% confidence level

Table 11A: Inside vs. Outsize Records as Predictor of Subsequent Performance

In the regression below I include (i) all new customers that reported no claims in the preceding three years, and (ii) all repeat customers that have been with the insurer during the preceding three years and had no claims.

Dependent variable:	**Total insurance payment**	**Total annual cost**	**Number of claims**
New	194.5^{***} (85.5)	316.0^{***} (88.05)	0.033^{***} (0.005)
Time fixed effect	YES	YES	YES
N	93,812	93,812	93,812
Adj-R^2	0.010	0.010	0.015

- ***, **, * - Significant at 1%, 5%, and 10% confidence level, respectively

Table 11B: Past Record of Departing and Staying Customers: Summary Statistics

	Company experience	1	2	3	4	5
	Deductible level	No claims	No claims	No claims	No claims	No claims
Group 1	Low-left	71.94	73.99	72.40	74.36	-------
	Low-stayed	79.27	81.09	82.02	85.60	91.03
	Regular-left	74.17	72.76	76.18	77.26	-------
	Regular-stayed	82.02	84.23	83.71	86.24	92.62
Group 2	Low-left	72.5	72.46	69.89	-------	
	Low-stayed	79.82	80.47	83.48	89.23	
	Regular-left	75.75	74.48	76.43	-------	
	Regular-stayed	83.58	84.17	85.85	92.53	
Group 3	Low-left	73.2	73.25	-------		
	Low-stayed	80.75	81.48	87.02		
	Regular-left	75.11	74.26	-------		
	Regular-stayed	82.29	85.36	90.85		
Group 4	Low-left	74.00	-------			
	Low-stayed	81.56	84.90			
	Regular-left	77.33	-------			
	Regular-stayed	83.95	91.17			

Table 11C: Past Record of Departing vs. Staying Customers

Dependent variable: stayed		
	OLS	**LOGIT**
Number of claims	-0.102***	-0.522***
	(0.002)	(0.0124)
Deductible level	0.0157***	0.1182***
	(0.002)	(0.016)
Time Fixed Effect	YES	YES
N	156,245	156,245
Adj-R^2	0.3100	0.3021

- I only report the coefficients of interest
- ***,**,* - Significant at 1%, 5%, and 10% confidence level, respectively

Table 11D: The Performance of Repeat Customers

Testing whether, within each group (low/regular deductible), the number of claims is lower for individuals with more years of experience with the insurer:

Dependent variable: number of claims

Deductible level:	**Low**		**Regular**	
	OLS	**std.**	**OLS**	**std.**
Buyer's demographic characteristics:				
Age	0.0015	0.001	0.0004	0.0004
Gender	0.036	0.033	-0.019*	0.011
Age-gender	-0.0007	0.0007	0.0004	0.0003
Single	-0.017	0.014	0.004	0.005
Academic education	0.002	0.009	-0.003	0.004
Buyer's car characteristics:				
CC	3.5e-06	1.5e-05	1.6e-06	4.7e-06
Car model year	-0.004**	0.002	-0.002**	0.0008
Value of car	3.2e-07	2.5e-07	2.43e-07***	7.9e-08
Social and private se	0.013	0.017	0.010*	0.006
Main car	0.004	0.012	-0.009**	0.004
Buyer's driving characteristics:				
Driving experience	-0.0007	0.0007	0.0005*	0.0003
Number of claims last year	0.043***	0.013	0.015**	0.006
Number of claims two years ago	0.026**	0.013	0.017**	0.007
Number of claims three years ago	0.024	0.015	0.002	0.010
Number of claims four years ago	0.027	0.021	0.065***	0.017
No experience last year	0.101*	0.059	-0.013	0.016
No experience two years ago	-0.059	0.064	0.018	0.018
No experience three years ago	0.014	0.042	0.013	0.011
Young driver age	-0.018	0.013	-0.008*	0.005
Young driver experience	0.009	0.017	-0.006	0.005
Young driver gender	-0.045	0.021	-0.024***	0.007
Young driver index	0.143***	0.045	0.111***	0.015
Claim1	0.027***	0.016	0.029***	0.005
Claim2	0.022	0.018	0.019***	0.006
Claim3	0.065***	0.013	0.032***	0.007
Number of drivers	0.016	0.011	-0.002	0.003
Other:				
Company years of experience	**-0.013*** **	**0.003**	**-0.003*** **	**0.001**
Period	0.121***	0.015	0.120***	0.005
Constant	7.706	4.748	3.943**	1.554
N	7,712		49,698	
Adj-R^2	0.0167		0.0172	

- This regression includes only data for the company fifth year of operation. Doing the same for the whole sample (for the whole five years of the company operation) yields similar results.
- ***, **, * - Significant at 1%, 5%, and 10% confidence level, respectively

Table 12: Profits on New and Repeated Customers

Dependent variable:	**LRETAC**
	OLS
Company years of experience	-0.0094***
	(0.0003)
N	213,642
Adj-R^2	0.031

- I only report the coefficient of interest
- All the coefficients are significant with 1% confidence level

Figure 1

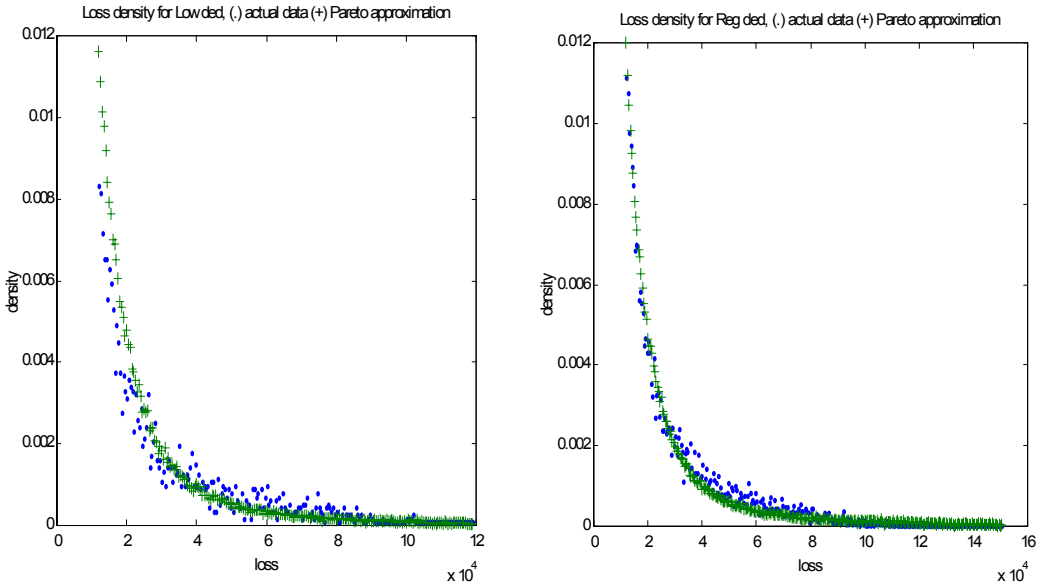
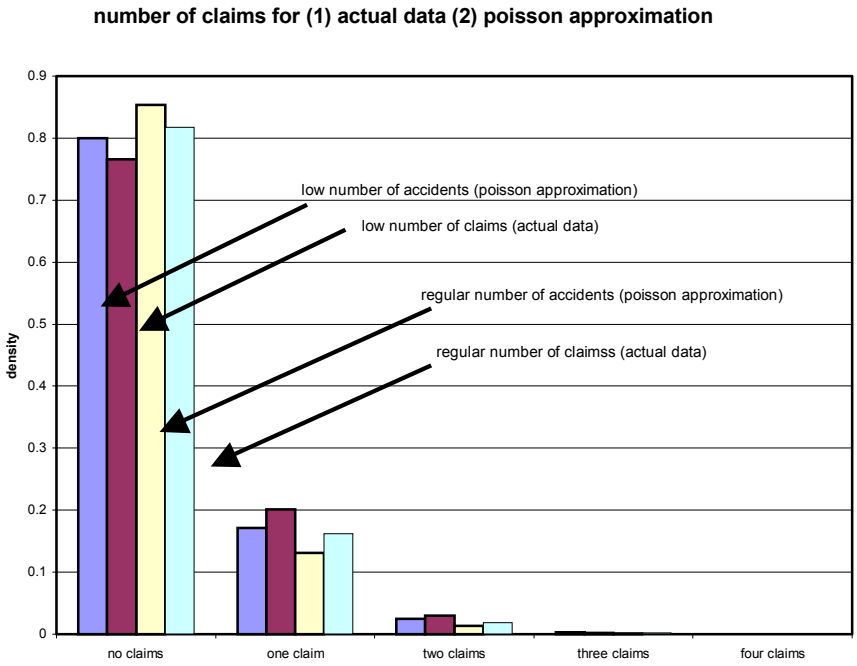


Figure 2



Number of accidents/claims	For Low Density for accidents	For Low Density for claims	For Regular Density accidents	For Regular Density claims
0	0.799	0.766	0.854	0.817
1	0.172	0.201	0.131	0.162
2	0.025	0.030	0.014	0.018
3	0.003	0.003	0.002	0.002
4	0.0004	0.0002	0.0001	0.0001