NBER WORKING PAPER SERIES

ESTIMATING RISK PREFERENCES FROM DEDUCTIBLE CHOICE

Alma Cohen
Liran Einav

Working Paper 11461
http://www.nber.org/papers/w11461

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2005

We thank Susan Athey, Lanier Benkard, Tim Bresnahan, Amy Finkelstein, Igal Hendel, Mark Israel, Felix Kubler, Jon Levin, Aviv Nevo, Harry Paarsch, Peter Rossi, Esteban Rossi-Hansberg, Morten Sorensen, Frank Wolak, and seminar participants at Chicago GSB, Duke, Haas, Harvard, Minnesota, Northwestern, Penn State, Stanford, UW Madison, SITE 2004, the Econometric Society 2005 Winter meeting, and the NBER IO 2005 Winter meeting for helpful comments and suggestions. All remaining errors are, of course, ours. Financial support from the National Science Foundation (Grant No. SES-0452555) and from Stanford's Office of Technology Licensing is gratefully acknowledged. Address correspondence to Liran Einav, Department of Economics, Stanford University, Stanford, CA 94305-6072; Tel: (650) 723-3704; E-mail: leinav@stanford.edu. Cohen affiliations: Analysis Group L.L.C. and John M. Olin Center for Law, Economics, and Business at Harvard Law School. E-mail: acohen@analysisgroup.com Einav affiliations: Department of Economics, Stanford University and NBER. E-mail: leinav@stanford.edu. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Alma Cohen and Liran Einav. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating Risk Preferences from Deductible Choice
Alma Cohen and Liran Einav
NBER Working Paper No. 11461
June 2005
JEL No. D82, G22

ABSTRACT

We use a large data set of deductible choices in auto insurance contracts to estimate the distribution of risk preferences in our sample. To do so, we develop a structural econometric model, which accounts for adverse selection by allowing for unobserved heterogeneity in both risk (probability of an accident) and risk aversion. Ex-post claim information separately identifies the marginal distribution of risk, while the joint distribution of risk and risk aversion is identified by the deductible choice. We find that individuals in our sample have on average an estimated absolute risk aversion which is higher than other estimates found in the literature. Using annual income as a measure of wealth, we find an average two-digit coefficient of relative risk aversion. We also find that women tend to be more risk averse than men, that proxies for income and wealth are positively related to absolute risk aversion, that unobserved heterogeneity in risk preferences is higher relative to that of risk, and that unobserved risk is positively correlated with unobserved risk aversion. Finally, we use our results for counterfactual exercises that assess the profitability of insurance contracts under various assumptions.

Alma Cohen
Analysis Group
111 Huntington Avenue, 10th Floor
Boston, MA 02199
acohen@analysisgroup.com

Liran Einav
Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072
and NBER
leinav@stanford.edu

1 Introduction

The analysis of decisions under uncertainty is central to economics. Indeed, expected utility theory, which is typically used to explain such decisions through the use of risk aversion, is one of the discipline's most celebrated theories.[1] But how risk averse are individuals? How heterogeneous are individuals' attitudes towards risk? How do they vary with individuals' characteristics? These are all questions, which, somewhat surprisingly, have received only little attention in empirical microeconomics.

Much of the existing evidence about risk preferences is based on introspection, laboratory experiments,[2] data on bettors or television game show participants,[3] answers given by individuals to hypothetical survey questions,[4] and estimates that are driven by the imposed functional-form relationship between static risk taking behavior and inter-temporal substitution.[5] We believe that supporting these findings using direct evidence from risky decisions made by actual market participants is important.

In this study we estimate risk preferences from the choice of deductible in auto insurance contracts. In particular, we exploit a rich data set of more than 100,000 individuals choosing from an individual-specific menu of four deductible-premium combinations offered by an Israeli auto insurance company. An individual who chooses low deductible is exposed to less risk, but is faced with higher level of expected expenditure. Thus, the decision to choose the low (high) deductible provides a lower (upper) bound for the coefficient of (absolute) risk aversion for each individual. Variation in the deductible-premium choices available to each individual in our data allows us to identify the distribution of the attitudes towards risk in our sample.

Obtaining measures of risk aversion from participants in insurance markets is particularly important, as risk aversion is the primary reason for the existence of such markets. Thus, measuring risk aversion and its interaction with risk will have direct and important implications for profitability of market participants, for market efficiency, and for potential policy interventions. Therefore, to the extent that it may not be straightforward to extrapolate utility parameters from one market context to another, it seems useful to obtain such parameters from the same markets to which they are subsequently applied.

In our view, the deductible choice is (almost) an ideal setting for estimating risk aversion from insurance data. Other decisions in an insurance context may involve additional preference-based explanations for coverage choice, which are unrelated to financial risk and will make inference about

[1] Notwhistanding the recent debate about the empirical relevance of expected utility theory (Rabin, 2000; and Rabin and Thaler, 2001), which is discussed later.

[2] See, for example, Kachelmeier and Shehata (1992), Smith and Walker (1993), and Holt and Laury (2002).

[3] See Gertner (1993), Metrick (1995), Jullien and Salanie (2000), and Beetsma and Schotman (2001).

[4] See, for example, Viscusi and Evans (1990), Evans and Viscusi (1991), Barsky et al. (1997), Donkers et al. (2001), and Hartog et al. (2002) .

[5] Much of the finance and macroeconomics literature, going back to Friend and Blume (1975), relies on this assumption. As noted by Kocherlakota (1996) in a review of this literature, the level of static risk aversion is still a fairly open question.

risk aversion more difficult. This is the case for the choice among health plans, annuities, or just whether to insure or not.[6] In contrast, the choice among different alternatives that vary *only* in their financial parameters (namely the levels of deductibles and premia) is a case in which the effect of risk aversion can be more plausibly isolated and estimated.

We are aware of only few attempts to recover risk preferences from decisions of regular market participants. Saha (1997) looks at firms' production decisions, and Chetty (2004) recovers risk preferences from labor supply decisions. The important study by Cicchetti and Dubin (1994) is probably the closest work to ours. They look at individuals' decisions whether or not to insure against failure of inside telephone wires. The cost of the insurance they analyze is 45 cents per month, and it covers a damage of about 55 dollars, which occurs with probability of 0.005 per month. In our auto insurance data events are more frequent and commonly observed, stakes are higher, the potential loss (the difference between the deductible amounts) is known, and the choice we analyze is more immune to alternative preference-based explanations. Our paper also differs in its methodology; in particular, we allow for unobserved heterogeneity in risk preferences, while Cicchetti and Dubin (1994) do not.[7]

Heterogeneity in risk preferences is not, of course, the only possible explanation for deductible choices. Heterogeneity in accident risk (adverse selection) should also be accounted for. The average deductible-premium menu in our data offers an individual to pay an additional premium of \$55 in order to save \$182 in deductible payments, in the event of a claim. This menu implies that a risk-neutral individual should choose higher coverage (low deductible) if and only if her claim propensity is $\frac{55}{182} = 30\%$ or more. As the average annual claim propensity for individuals in our data is 24.5%, this additional coverage is actuarially unfair. Despite this, about 18% of the individuals we observe choose to pay for the lower deductible. Within the expected utility framework, and abstracting from moral hazard (see later), this can happen for one of two reasons: these individuals have either higher risk exposure (claim propensity) than the average individual, or have higher risk aversion, or both. Ex-post claim information is used to identify between these two possibilities, and to estimate the joint distribution of risk and risk aversion.[8]

Two important aspects of the data should be noted. First, we observe all the variables that are observed by the insurance company. Therefore, at least in principle, once we condition on observables, any remaining variation in the deductible-premium menus offered to consumers is econometrically exogenous: it cannot depend on unobserved characteristics of consumers. Second,

[6] For example, Rabin and Thaler (2001, footnote 2) point out that one of their colleagues buys the insurance analyzed by Cicchetti and Dubin (1994) in order to improve the service he will get in the event of a claim. We think that our deductible choice analysis is immune to such critic.

[7] The possibility of using deductibles to make inference about risk aversion was first pointed out by Dreze (1981), who suggested a different strategy from the one we employ. While we rely on individuals' choices of deductibles ("demand side" information), Dreze (1981) suggested a method that relies on the optimality of the observed contracts ("supply side" information).

[8] Through most of the paper we assume that individuals perfectly know their risk types. Later in the paper we also consider heterogeneity in the information individuals have regarding their risk types, and show that the qualitative results of the paper are not very sensitive to the complete information assumption.

we have data on the ex-post realization of the number of claims for each policy. This helps us identify the model; it allows us to estimate the joint distribution of risk and risk aversion, thereby accounting for adverse selection in the choice of deductibles. Loosely speaking, the distribution of claims identifies the marginal distribution of risk types. This distribution allows us to compute the posterior distribution (conditional on the number of claims) of risk types for each individual, and to integrate over it when we analyze the individual's deductible choice. This accounts for adverse selection, as individuals with more claims will have a less favorable posterior risk distribution. Such distribution will make these individuals more likely to choose higher coverage (lower deductible), even in the absence of heterogeneity in risk aversion.

The majority of the existing adverse selection literature addresses the important question of whether adverse selection exists in different insurance markets. As suggested by the influential work of Chiappori and Salanie (2000), it uses "reduced form" specifications to test whether, after controlling for observables, outcomes and coverage choices are significantly correlated.[9] As our main goal is quite different, we take a more structural approach. By assuming a particular structure for the adverse selection mechanism, we can account for it when estimating the distribution in risk preferences, which is the main objective of the paper. Moreover, while the structure of adverse selection is assumed, its existence and relative importance are not imposed; the structural assumptions allow us to estimate the importance of adverse selection relative to the selection induced by unobserved heterogeneity in risk attitudes.[10]

The structural assumptions also allow us to estimate the correlation between risk type and risk aversion. Recently, it has been argued[11] that a negative correlation between risk aversion and risk types may be the reason why several other important studies (for example, Chiappori and Salanie, 2000) did not find empirical evidence for adverse selection in insurance markets. Our analysis provides direct evidence about this important relationship; our estimate of strong positive correlation between risk and risk aversion suggests that, at least in our data, ignoring heterogeneity in risk preferences will go in the other way: it will make us even more likely to find evidence for adverse selection.[12]

We make two important assumptions throughout the paper. First, we assume that by choosing the low deductible, in the event of an accident the individual "gains" the difference between the high and the low deductible. This is not completely true, as with some probability the amount of the claim would fall between these two deductible levels, so the individual should, in principle, take into account the loss distribution when it falls in this range. Our data include, however,

[9] See also Puelz and Snow (1994), Finkelstein and McGarry (2003), and Finkelstein and Poterba (2004), as well as Cohen (2005) for our data.

[10] This approach is somewhat similar to that of Cardon and Hendel (2001). In our discussion of identification (Section 3.3) we discuss the conceptual similarities and differences in more detail.

[11] See De Meza and Webb (2001), Finkelstein and McGarry (2003), Jullien, Salanie, and Salanie (2003), and Israel (2005).

[12] Throughout the paper we use the term "adverse selection" to denote selection on risk, while selection on risk aversion is just "selection." Some of the literature calls both of these selection mechanisms "adverse selection," with the distinction being between "common values" (selection on risk) and "private values" (selection on risk aversion).

the amount of the claim, and analyzing the claim amounts of those individuals who chose low deductibles suggests that only one percent of the claims made would not have been worthy to file with the higher deductible level, thereby making our assumption not very restrictive. This is true even when taking into account the (small, in Israel) dynamic costs of filing a claim through its effect on experience rating. The second assumption we make is to abstract from moral hazard. By introspection we do not think that moral hazard plays an important role in this setting.[13] While behavior may be affected by the decision whether to insure or not, the deductible choice is less likely to affect behavior. Furthermore, if moral hazard exists, our data cannot separately identify it from adverse selection unless we impose additional structural restrictions. Moral hazard can be separately identified only if one observed the behavior of the same individual after (exogenously) making different contract choices.[14] Finally, one should note that the introduction of moral hazard to the setup is likely to reduce the attractiveness of a low deductible, thereby biasing our estimates of risk aversion downwards. In that case, our estimate can be viewed as a lower bound on the true level of risk aversion. These assumptions and their implications are discussed in more detail in Section 5.

In our benchmark model we assume that individuals have full information about both their risk exposure and their level of risk aversion, and that claims are the realization of a Poisson process with individual-specific claim rate.[15] The assumption that claims are generated by a one-parameter distribution is crucial, as it allows us to uniquely back out the distribution of risk types from only claim data. This assumption facilitates the identification of unobserved heterogeneity in risk preferences. For computational convenience, we also assume that the joint distribution of risk type and the coefficient of absolute risk aversion follows a bivariate Lognormal distribution. We use expected utility theory to characterize the deductible choice as a function of risk and risk aversion, imposing minimal assumptions on the particular functional form of the vNM utility function. Thus, conditional on the menu of deductible-premium combinations and on the (unobserved) individual-specific risk, the deductible choice can be estimated using a simple Probit. While this set up is somewhat similar to the first-stage regression of a selection model (Heckman, 1979), there are two important differences which make our empirical model more complicated. First, we do not observe claim rate directly, but only a random realization of it. Thus, we need to integrate over the posterior distribution of claim rates, conditional on the observed realization. Second, the claim rate enters the "first stage equation" not only through the correlation between the error terms, as in a typical selection model, but also directly, as a result of adverse selection.[16] These differences make the estimation of our model by standard Likelihood techniques (or, alternatively, GMM) less attractive,

[13] This assumption is also consistent with Cohen and Einav (2003), who find no evidence for behavioral response to changes in seat belt laws.

[14] See also Chiaporri and Heckman (2000).

[15] The Poisson assumption may be violated if one is concerned about a dynamic change in incentives after the first accident. In such a case, if the timing of claims were available, one could replace the outcome variable by the time until the first accident, and repeat the same exercise.

[16] This structural component of the direct effect is also what identifies the level of risk aversion.

as they would require us to separately compute unattractive integrals for each individual in our data, and for each value of the parameters. Instead, we use Bayesian econometrics by applying Markov Chain Monte Carlo techniques and estimating the model using a Gibbs Sampler. This approach only requires us to sample from truncated univariate Normal distributions, significantly reducing the computational burden.

Our estimates of the level of risk aversion are on average significantly greater than other estimates in the literature. The mean level of the coefficient of absolute risk aversion (translated to dollar amounts) is 0.0016. This implies that a quadratic expected utility maximizer will be indifferent about participating in a lottery in which she gains $100 with probability 0.5 and loses $61.3 with probability 0.5. Imposing a CARA expected utility function significantly reduces this estimate, but still keeps it at higher risk aversion levels than those estimated in the literature. For example, the parameter estimates of Gertner (1993) and Metrick (1995) suggest that their representative TV show participant will be indifferent about participating in a 50-50 lottery of gain $100 lose $97 and gain $100 lose $99.3, respectively. Using the average annual income (in Israel) as a measure of wealth, we find an average two-digit coefficient of relative risk aversion, which, again, is much higher than other estimates present in the literature. We also find that risk aversion does not significantly change with age, that females are, on average, more risk averse than males, and that different proxies for higher income or wealth are associated with higher levels of risk aversion.

Turning to the joint distribution of risk and risk aversion, we have two important findings. First, we find that the unobserved heterogeneity in risk aversion is relatively much higher than that in risk exposure. This finding reduces the importance of accounting for adverse selection in the design of optimal auto insurance contracts. Second, conditional on observables, we find that unobserved risk has a strong positive correlation with unobserved risk aversion. This correlation is driven by the fact that the number of claims and the propensity to choose higher coverage (low deductible) are highly correlated in the data. Given our structural assumptions, the observed correlation is too high to be explained only by adverse selection. The only other way the model can explain such a strong correlation is through positive correlation between risk and risk aversion. We realize, of course, that this positive correlation is somewhat counterintuitive. It is natural to speculate that risk attitudes towards financial decisions are related to risk attitudes that affect driving behavior. This by itself should lead to negative relationship between risk aversion and claim propensity. We should note, however, that there are many other factors that affect claim propensities, which may go in the other direction. For example, wealthier people may be less risk averse and, at the same time, have lower claim propensities due to, say, shorter commute. We do not observe individual wealth or income, so such omitted factors may generate the positive correlation we find. More generally, we should note that as the measure of risk is very different from one market context to another, we do not think that these last findings about the joint distribution should necessarily generalize to other insurance markets. One way in which auto insurance may be special is that extremely cautious drivers may, in fact, expose themselves to greater risk. This is unlikely to happen in, say, health or life insurance markets.

Our empirical strategy and results may also help in guiding the recent theoretical literature on multi-dimensional screening. Our model presents two dimensions of unobserved heterogeneity, while the insurer has only a one-dimensional instrument for screening (price). This case is different from multi-dimensional screening in which the number of instruments is equal to the dimensions in which individuals differ.[17] Therefore, by construction, optimal contracts will necessitate some degree of "bunching".[18] The optimal shape of the contracts, however, will crucially depend on the relative variance of the two dimensions, as well as on their correlation structure. In Section 4.3 we use our estimation results to illustrate this point. We analyze the estimated profits of the insurer from offering different sets of deductible-premium combinations. We estimate that by offering a menu of contracts, rather than a single deductible-premium alternative, the operating profits of the insurer are higher by about 0.35%. The results also suggest that these *additional* profits can almost double by re-optimizing and increasing the level of the low deductible.

Much of our analysis focuses on estimating the *absolute* level of risk aversion. After all, this is what is identified by the choices made in our data, and this is what is relevant for analyzing the effects of alternative pricing policies in the auto insurance market. Any claim about *relative* risk aversion must employ additional assumptions about individuals' wealth. This is for two reasons. First, we do not directly observe individuals' wealth, so we will need to make assumptions about how wealth is related to the variables we do observe. Second, much of the recent debate about the empirical relevance of expected utility theory has focused on identifying the relevant wealth one uses in different contexts.[19] Our exercise is neutral with respect to either side of the debate. In our view, the debate focuses on how the curvature of the vNM utility function varies with wealth, or across different settings. We only measure the curvature of the vNM utility function at a particular wealth level, whatever this wealth level may be. By allowing unobserved heterogeneity in this curvature across individuals, we place no conceptual restrictions on the relationship between wealth and risk aversion. Our estimated distribution of risk preferences can be thought of as a convolution of the distribution of (relevant) wealth and risk attitudes. We do not attempt to break down the distribution into these two components.

More generally, one should view our two-dimensional space of risk type and risk aversion in the context of the variables we use to identify it. In particular, we use claim data to identify risk type, thereby leaving everything else to be interpreted as risk aversion. Thus, for example, overconfidence will be captured by a lower level of estimated risk aversion. At the conceptual level, we do not think that this should be viewed as a problem, as long as the results are interpreted in the right way.[20] The interpretation may be more important once our results are extrapolated to

[17] Randomized insurance contracts (Arnott and Stiglitz, 1988; Landsberger and Meilijson, 1999) may provide a second dimension to screen high risk aversion individuals from high risk. Such contracts, however, are not practical in many markets.

[18] See Landsberger and Meilijson (1999), Smart (2000), Jullien, Salanie, and Salanie (2003), and Villeneueve (2003) for some related theoretical results. From a theory standpoint, the tools provided by Armstrong (1999) will most likely be useful for such analysis.

[19] See Rabin (2000), Rabin and Thaler (2001), Rubinstein (2001), Watt (2002), and Palacio-Huerta et al. (2003).

[20] In practice, of course, our particular interpretation leads us to impose a particular trade-off between risk and risk aversion. Other interpretations would lead to similar, albeit different, functional forms.

other environments, in which those behavioral biases may potentially take a different form.[21] We discuss this further in the concluding section. Finally, individuals may, of course, vary in other dimensions. While any attempt to identify additional dimensions of heterogeneity will necessarily rely on structural assumptions, we perform one important exercise as a robustness check of our main results. In this exercise, we relax the assumption that individuals perfectly know their risk types. Rather, they learn their risk type through a simple Bayesian learning model. This model allows heterogeneity in the extent to which individuals know their own risk type at the time of the deductible choice. The qualitative results remain the same. In Section 5 we discuss in more detail the sensitivity of the results to this and other restrictions imposed in the benchmark model.

The paper continues as follows. Section 2 describes the data. Section 3 lays out the structural model, describes the estimation strategy, and discusses the identification of the model. Section 4 describes the results and performs several robustness tests and counterfactual experiments. Section 5 provides an informal discussion of the sensitivity of the results to various key assumptions, and Section 6 concludes.

2 Data

2.1 General Description

We use data obtained from an insurance company that operates in the market for automobile insurance in Israel. The data contain information about *all* 105,800 new policyholders who joined the insurer and purchased (annual) policies from it during the first five years of the company's operation, from November 1994 to October 1999. Although many of these individuals stayed with the insurer in subsequent years, we restrict attention to the deductible choices made by each individual in her *first* contract with the company. This allows us to abstract from the selection implied by the endogenous choice of individuals whether to remain with the company or not (see Cohen, 2003 and 2005).

The company studied was the first company in the Israeli market that marketed insurance to customers directly rather than through insurance agents. By the end of the studied period, the company sold about seven percent of the automobile insurance policies issued in Israel. Direct insurers operate in many countries including the US and appear to have a significant cost advantage (Cummins and Van Derhei, 1979). The studied company estimated that selling insurance directly results in a cost advantage of 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 (D'Arcy and Doherty, 1990). This is so because their product involves the "disamenity" of not having an agent to work with and turn to. The costs of this disamenity appear to be substantial for a large fraction of potential buyers of insurance. Consequently, only a subset of the insurance buyers are open to considering buying insurance

[21] For example, one may speculate that over confidence is more of an issue in auto insurance than it is in life insurance.

directly and thus constitute potential customers of direct insurers. This aspect of the company clearly makes the results of the paper applicable only to those consumers who are willing to buy direct insurance; Section 5 discusses this selection in more detail.

While the paper's primary focus is on the "demand side" of the market, namely on the deductible choices, it is important to think about the "supply side" of the market (pricing), as this will be relevant for any counterfactual exercise as well as for understanding the viability of the outside option (which we do not observe). In this context, one should note that, for the most part, the company had substantial market power over its pool of customers. This makes monopolistic screening models apply more naturally than competitive models of insurance (e.g. Rothschild and Stiglitz, 1976). In particular, during the first two years of the company's operations, the prices it offered were considerably lower, by roughly 20%, than those offered by other, "conventional" insurers. In the company's third year of operation (December 1996 to March 1998) it faced more competitive conditions. During this year, the established companies, trying to fight off the new entrant, lowered the premia for policies with regular deductibles to the levels offered by the company. In the remaining period included in the data, the established companies raised their prices back to previous levels, leaving the company again with a substantial price advantage.[22]

For each policy, our data set includes *all* the information that the insurer had about the characteristics of the policyholder: the policyholder's demographic characteristics, the policyholder's vehicle characteristics, and the policyholder's driving experience characteristics. The appendix provides a list of the variables with precise definitions, and Table 1 provides descriptive statistics. In addition, our data set includes the individual-specific menu of four deductible-premium combinations that the individual was offered (see later), the individual's choice from this menu, and the realization of risks covered by the policy: the length of the period over which it was in effect, the number of claims submitted by the policyholder, and the amounts of the submitted claims that the insurer paid or was expected to pay.[23] Finally, we use the zip codes of the policyholders' home addresses to augment the data with proxies for additional individual characteristics based on the Israeli 1995 census.[24] In particular, the Israeli Central Bureau of Statistics (CBS) associates each census respondent with a unique "statistical area" (census tract), each including between 1,000 and 10,000 (relatively homogeneous) residents. We matched these census tracts with zip codes based on street addresses, and constructed zip code level variables. These constructed variables are available for more than 80% of the policyholders. The most important such variable is that of (gross) monthly income, which is based on self-reported income by census respondents augmented (by the CBS) with social security data.

[22] During this last period, two other companies offering insurance directly were established. Due to first-mover advantage (as viewed by the company's management), which helped the company maintain strong position in the market, these two new companies did not affect much pricing policies until the end of our observation period. Right in the end of this period the studied company acquired one of those entrants.

[23] Throughout the analysis, we make the assumption that the main policyholder is the individual who makes the deductible choice. Clearly, to the extent that this is not always the case, the results should be interpreted accordingly.

[24] The company has the addresses on record for billing purposes. Although, in principle, the company could have used these data for pricing, they do not do so.

The policies offered by the insurer (as are all policies offered in the Israeli automobile insurance market) are one-period policies with no commitment on the part of either the insurer or the policyholder.[25] The auto-insurance policy we analyze in this paper resembles the US version of "comprehensive" insurance, but is not exactly the same. It is not mandatory, but it is believed to be held by a large fraction of Israeli car owners (above 70%, according to the company's executives; we are not aware of any data about this). The policy does not cover death or injuries to the policyholder or to third parties, which are insured through a separate, mandatory policy. It is worth noting that the deductible choice is irrelevant for certain types of damages covered by the policy. Insurance policies for radio, windshield, replacement car, and towing services are structured and priced separately. Auto theft, total-loss accidents, and not "at fault" accidents are covered by the policy, but do not involve deductible payments. Accordingly, we record as a claim only those claims that eventually resulted in deductible payments. These are the only claims relevant for the deductible choice.

Throughout the paper, we use and report money amounts in current (nominal) New Israeli Shekels (NIS) to avoid creating artificial variation in the data. Consequently, the following facts may be useful for interpretation and comparison with other papers in the literature. The exchange rate between NIS and US dollars monotonically increased from 3.01 in 1995 to 4.14 in 1999 (on average, it was 3.52).[26] Annual inflation was about eight percent on average, and cumulative inflation over the observation period was 48%. We will account for these effects, as well as other general trends, by using year dummy variables throughout the analysis.

2.2 The Menu of Deductible-Premium Combinations

Denote by x_i the vector of characteristics individual i reports to the insurance company, as described above. After learning x_i the insurer offered individual i a menu of four contract choices. One option, which was labeled "regular" by the company, offered a "regular" deductible and a "regular" premium. The term "regular" was used for this deductible level both because it was relatively similar to the deductible levels offered by other insurers and because most policyholders (about 80%, see Table 2A) chose it. The regular premium varied across individuals according to some deterministic function (unknown to us), $p_{it} = f_t(x_i)$, which was quite stable over time. The premia associated with the other options were computed by multiplying p_{it} by three different constants, as described in the first row of the table below. Similarly, the regular deductible, d_{it}, was converted to three other offered deductible levels using three other constants (see the table below). The regular deductible level was directly linked to the regular premium level through $d_{it} = \min\{\frac{1}{2}p_{it}, cap_t\}$.

[25] There is a substantial literature that studies the optimal design of policies that commit customers to a multi-period contract, or that include a one-sided commitment of the insurer to offer the policyholder certain terms in subsequent periods (Dionne and Lasserre, 1985; Cooper and Hayes, 1987; Dionne and Doherty, 1994; Hendel and Lizzeri, 2003). Although such policies are observed in certain countries (Dionne and Vanasse, 1992), many insurance markets, including the Israeli one we study, use only one-period no-commitment policies (Kunreuther and Pauly, 1985).

[26] PPP figures, which may be more relevant for comparison, were about 10% lower than the nominal exchange rates, running from 2.60 in 1995 to 3.74 in 1999.

Namely, for regular premia which were not too high, the level of the regular deductible was set at 50% of the (regular) premium. For higher regular premia, the regular deductible level was set at a uniform cap that varied over time but not across individuals.

The Pricing Formula

	"Low"	"Regular"	"High"	"Very High"
Premium	$1.06 \cdot p_{it}$	p_{it}	$0.875 \cdot p_{it}$	$0.8 \cdot p_{it}$
Deductible	$0.6 \cdot d_{it}$	d_{it}	$1.8 \cdot d_{it}$	$2.6 \cdot d_{it}$

To understand the pricing formula, suppose that for a given individual the company's formula yielded a "regular" premium of 2,000 NIS. Suppose that the deductible cap was set at the time at 1,500 NIS, so it was not binding for this particular individual. This implies that the "regular" premium-deductible contract offered to her was (2,000; 1,000). Consequently, the "low," "high," and "very high" contracts were set at (2,120; 600), (1,750; 1,800), and (1,600; 2,600), respectively. A different individual who was quoted contract terms at the same time, but had characteristics which yielded a higher "regular" premium of 4,000 NIS, had the uniform cap binding, and was quoted the following premium-deductible combinations: {(4,000; 1,500), (4,240; 900), (3,500; 2,700), and (3,200; 3,900)} for "regular," "low," "high," and "very high," respectively.

There are two main sources of what we view as exogenous variation, resulting from company's experimentation and discrete adjustments to inflation and competitive conditions. The first source of variation arises from variation in the multipliers used to construct the menu of contracts. While the multipliers described above were fixed across individuals and over time, there was a six month period during the insurer's first year of operation (May 1995 to October 1995), in which the insurer experimented with multipliers which were slightly modified. For individuals with low levels of regular premia during the specified period, the regular deductible was set at 53% (instead of 50%) of the regular premium, the low deductible was set at 33% (instead of 30%) of the regular premium, and so on. This modified formula covers almost ten percent of the sample. The second source of variation in the menus offered arises from changes over time to the uniform cap, cap_t. The cap varied over time (in both directions, up and down) due to inflation, competitive conditions, and as the company gained more experience. Figure 1 presents the way the uniform cap changed over our observation period. The cap was binding for about a third of the policyholders in our data. One should also note that a small change in the cap does not only affect the menus offered to those potential customers who move from being above the cap to below the cap; the change affects *all* potential customers whose previous regular premia were higher than the cap, as all their menus will stipulate higher levels of deductibles. Figure 2 plots the (unconditional) variation of menus in the data. As can be seen, much of this variation is driven by the exogenous shifts in the uniform deductible cap. The underlying assumption is that, conditional on observables, these sources of variation primarily affect the deductible choice of new customers, but they do not have a significant impact on the probability of purchasing insurance from the company. This assumption holds in the data with respect to observables.

2.3 Descriptive Figures

The top part of Table 2A provides descriptive statistics for the deductible-premium menus offered, all of which calculated according to the formula described above. Only one percent of the policyholders chose the "high" or the "very high" deductible options. Therefore, for the rest of the analysis we only focus on the choice between the two other options: "regular" (chosen by eighty percent) and "low" (chosen by almost twenty percent). The small frequency of "high" or "very high" choices provides important information about the lower ends of the risk and risk aversion distributions, but (for that same reason) makes the analysis sensitive to functional form. Given these low frequencies, it is also unclear to us whether these options were frequently mentioned during the insurance sale process, rendering their use somewhat inappropriate.[27] We should emphasize that focusing only on the low and regular deductible levels does not create any selection bias because we do not omit the individuals who chose "high" or "very high" deductibles. For these individuals, we assume that they chose a "regular" deductible. This assumption is consistent with the structural model we develop in the next section: conditional on choosing "high" or "very high" deductibles, the individual would almost always prefer the "regular" over the "low" deductible.[28]

The bottom part of Table 2A, as well as Table 2B, presents some statistics for the realizations of the policies. We focus only on claim rates and not on the amounts of the claims. This is because any amount above the higher deductible level is covered irrespective of the deductible choice, and the vast majority of the claims fit in this category (see Section 5). For all these claims, the gain from choosing a low deductible is the same in the event of a claim; it is equal to the difference between the two deductible levels, irrespective of the claim amount. Thus, the claim amount is rarely relevant for the deductible choice decision (and, likewise, for the company's pricing decision we analyze in Section 4.3).

Averaging across all individuals, the annual claim rate was about twenty five percent. One can clearly observe some initial evidence of adverse selection: on average, individuals who chose the low deductible had higher claim rates (30.9%) than those who chose the regular deductible (23.2%). Those who chose high and very high deductibles had even lower claim rates (12.8% and 13.3%, respectively). It may be worth interpreting these figures in the context of the pricing formula described above. A risk neutral individual will choose the low deductible if and only if her claim rate is higher than $\frac{\Delta p}{\Delta d} = \frac{p^{low} - p^{regular}}{d^{regular} - d^{low}}$. When the deductible cap does not bind, which is the case for about two thirds of the sample, this ratio is directly given by the pricing formula and is equal to 30%. Thus, any individual with a claim rate higher than 30% will benefit from buying the additional coverage provided by a low deductible even without any risk aversion. The claim data suggest that the offered menu is cheaper than an actuarially fair contract for a non-negligible part of the population. This observation is in sharp contrast to other types of insurance contracts, such

[27] Considering these options creates a sharp lower bound on risk aversion for the majority of the observations, making the estimates much higher.

[28] This is always true for "very high" deductibles. For "high" deductibles, there is a small range of risk rates for which this is not true. Given the estimated coefficients, the probability of individuals falling within this region is less than one percent.

as appliance warranties, which are much more expensive than the actuarially fair price (Rabin and Thaler, 2001).

3 The Empirical Model

3.1 The Individual Decision Problem

Let w_i be individual i's wealth, (p_i^h, d_i^h) the insurance contract (premium and deductible, respectively) with high deductible, (p_i^l, d_i^l) the insurance contract with low deductible, t_i the duration of the policy, and $u_i(w)$ individual i's vNM utility function. We assume that the number of insurance claims is drawn from a Poisson distribution, with claims coming at a known (to the individual) rate of λ_i per unit of time (year). As already mentioned, we also assume that moral hazard does not play an important role, i.e. that λ_i is independent of the deductible choice and that, in the event of an accident, the value of the claim is greater than d_i^h with probability one. We discuss and justify these assumptions in Section 5. For the rest of this section, i subscripts are suppressed for convenience.

In the market we study, insurance policies are typically held for a full year, after which they can be automatically renewed with no commitment by either the company or the individual. Moreover, all auto-insurance policies sold in Israel can be canceled without prior notice by the policyholder, with premium payments being linearly prorated. It turns out to be convenient to think of the contract choice as a commitment for only a short amount of time, so both the premium and the probability of an accident (coming from a Poisson distribution) are proportional to the length of the time interval taken into account. This approach has several advantages. First, it helps to account for early cancellations and truncated policies (those which expired after October 1999, the end of our observation period), which together account for about thirty percent of the policies in our data.[29] Second, it makes the deductible choice independent of other longer-term uncertainties faced by the individual, allowing us to focus on the static risk-taking behavior, avoiding its relationship with inter-temporal substitution. Third, as will soon become clear, this formulation helps us to obtain a simple framework for analysis, which is attractive both analytically and computationally.[30]

[29] As can be seen in Table 2A, 70% of the policies in our data are observed through their full duration (one year). About 15% of the policies are truncated by the end of our observation period, and the remaining 15% are canceled for various reasons, such as change in car ownership, total-loss accident, or a unilateral decision of the policyholder to change insurance providers.

[30] This specification ignores the option value associated with not canceling a policy. This is not very restrictive. Since experience rating is small and menus do not change by much, this option value is likely to be close to zero. A simple alternative is to assume that individuals behave as if they commit for a full year of coverage. In such case, the model will be similar to the one we estimate, but will depend on the functional form of the vNM utility function, and would generally require taking infinite sums (over the potential realizations for the number of claims within the year). In the special case of quadratic expected utility maximizers, who only care about the mean and variance of the number of claims, this is easy to solve. The result is almost identical to the expression we subsequently derive in equation (7).

The expected utility that the individual obtains from the choice of a contract (p,d) is given by

$$\begin{aligned} v(p,d) &\equiv (1-\lambda t)u(w-pt)+(\lambda t)u(w-pt-d)= \\ &= u(w-pt)-\lambda t\left[u(w-pt)-u(w-pt-d)\right] \end{aligned} \tag{1}$$

We search for the individual who is indifferent between the two choices she is offered. This gives us a lower (upper) bound on the level of risk aversion for individuals who choose the low (high) deductible. Thus, we analyze the equation $v(p^h,d^h)=v(p^l,d^l)$, i.e.

$$u(w-p^ht)-\lambda t\left[u(w-p^ht)-u(w-p^ht-d^h)\right]=u(w-p^lt)-\lambda t\left[u(w-p^lt)-u(w-p^lt-d^l)\right] \tag{2}$$

By taking limits with respect to t (and applying L'Hopital's rule), we obtain

$$\begin{aligned} \lambda &= \lim_{t\to 0}\left(\frac{u(w-p^ht)-u(w-p^lt)}{t\cdot\left[(u(w-p^ht)-u(w-p^ht-d^h))-(u(w-p^lt)-u(w-p^lt-d^l))\right]}\right)= \\ &= \frac{(p^l-p^h)u'(w)}{u(w-d^l)-u(w-d^h)} \end{aligned} \tag{3}$$

or

$$(p^l-p^h)u'(w)=\lambda\left(u(w-d^l)-u(w-d^h)\right) \tag{4}$$

The last expression has a simple intuition. The right hand side is the expected gain (in utils) per unit of time from choosing a low deductible. The left hand side is the cost of such a choice per unit of time. For the individual to be indifferent, the expected gains must equal the costs.

We can now continue in one of two ways. In our benchmark specification, we try to avoid making functional form restrictions on the vNM utility function. By assuming that the third derivative of the vNM utility function is not too large, we can use a Taylor expansion for both terms on the right hand side of equation (4), i.e. $u(w-d)\approx u(w)-du'(w)+\frac{d^2}{2}u''(w)$. This gives us

$$\frac{p^l-p^h}{\lambda}u'(w)\approx(d^h-d^l)u'(w)-\frac{1}{2}(d^h-d^l)(d^h+d^l)u''(w) \tag{5}$$

Let $\Delta d\equiv d^h-d^l>0$, $\Delta p\equiv p^l-p^h>0$, and $\overline{d}\equiv\frac{1}{2}(d^h+d^l)$ to get

$$\frac{\Delta p}{\lambda\Delta d}u'(w)\approx u'(w)-\overline{d}u''(w) \tag{6}$$

or

$$r=\frac{-u''(w)}{u'(w)}\approx\frac{\frac{\Delta p}{\lambda\Delta d}-1}{\overline{d}} \tag{7}$$

where r is the coefficient of absolute risk aversion at wealth level w (recall, all notation is individual specific). The equation above defines an indifference set in the space of risk and risk aversion, which we will refer to by $(r_i^*(\lambda),\lambda)$ and $(\lambda_i^*(r),r)$ interchangeably. Note that both $r_i^*(\lambda)$ and $\lambda_i^*(r)$ have a closed-form representation, a property which will be computationally attractive for estimation.[31]

[31] For example, estimating the CARA version of the model, for which $r_i^*(\lambda)$ does not have a closed-form representation, takes almost ten times longer.

Note also that both terms are individual specific, as they depend on the deductible-choice menu, which varies across individuals.

Alternatively, we can impose a particular functional form on the vNM utility function. Two standard forms are those that exhibit constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA). In the CARA case, we can substitute $u(w) = -\exp(-rw)$ into equation (4) and rearrange to obtain

$$\lambda = \frac{r\Delta p}{\exp(rd^h) - \exp(rd^l)} \tag{8}$$

as the equation which defines the indifference set. Now, unlike before, there is no closed-form representation for $r_i^*(\lambda)$. Still, the set does not depend on wealth, which is a direct implication of the CARA property. Finally, in the CRRA case, we can substitute $u(w) = w^{1-\gamma}$ into equation (4) and rearrange to obtain

$$\lambda = \frac{(1-\gamma)w^{-\gamma}\Delta p}{(w-d^l)^{1-\gamma} - (w-d^h)^{1-\gamma}} \tag{9}$$

as the equation which defines the indifference set. To use this expression, we will also need to make assumptions about wealth, which we do not observe. The CARA and CRRA examples, and many other vNM utility functions, introduce a positive third derivative of $u(w)$, which provides an additional ("precautionary") incentive to insure. Therefore, in comparison to our benchmark specification which assumes a small, negligible third derivative, these specifications would lead to a greater incentive to choose a low deductible. In other words, in comparison to the benchmark specification and given λ we can use a lower level of absolute risk aversion to rationalize the low deductible choice. Thus, these specifications will generally lead to lower estimates of the coefficient of risk aversion.

For the rest of the paper, we regard each individual as associated with two-dimensional type parameters (r_i, λ_i), i.e. with her level of (absolute) risk aversion and with her level of risk. An individual with a risk level of λ_i, who is offered a menu $\{(p_i^h, d_i^h), (p_i^l, d_i^l)\}$ will choose the low deductible if and only if $r_i > r_i^*(\lambda_i)$. See Figure 3 for a graphical illustration.

3.2 The Benchmark Statistical Model

Our objective is to estimate the joint distribution of (λ_i, r_i) – the claim rate and coefficient of absolute risk aversion – in our population of policyholders, conditional on observables x_i. The benchmark formulation will impose that (λ_i, r_i) follows a bivariate Lognormal distribution.[32] Thus, we can write the model as

$$\ln \lambda_i = x_i'\beta + \varepsilon_i \tag{10}$$

$$\ln r_i = x_i'\gamma + \upsilon_i \tag{11}$$

[32] As will become clear later, the normality assumption provides a closed-form conditional distribution, allowing us to use only univariate (rather than bivariate) draws in the estimation procedure, significantly reducing the computational burden. There is hardly any literature on the distribution of risk preferences for us to draw from. The only evidence we are aware of is the experimental results presented by Andersen et al. (2004). These results show a skewed distribution with a fat right tail, which is qualitatively consistent with the Lognormal distribution.

with

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{pmatrix} \right) \tag{12}$$

The model becomes more complicated because neither λ_i nor r_i is directly observed. Thus, they are treated as latent variables, which, loosely speaking, can be thought of as random effects. We only observe two variables (the claims and the deductible choice) which are related to these two unobserved components. Thus, to complete our empirical model we have to specify the relationship between the observed variables and the latent ones. This is done by making two structural assumptions that were already mentioned. First, we assume that the number of claims is a realization from a Poisson distribution, namely

$$claims_i \sim Poisson(\lambda_i t_i) \tag{13}$$

where t_i is the observed duration of the policy. Second, we assume that individuals follow the theoretical model described in the previous section when they make the deductible choice. The model implies that individual i chooses low deductible ($choice_i = 1$) if and only if $r_i > r_i^*(\lambda_i)$, where $r_i^*(\cdot)$, defined in equation (7), has an individual-specific subscript because each individual faces a different deductible-premium menu. Thus, the empirical model for deductible choice is given by

$$\Pr(choice_i = 1) = \Pr(r_i > r_i^*(\lambda_i)) = \Pr\left(\exp(x_i'\gamma + \upsilon_i) > \frac{\frac{\Delta p_i}{\lambda_i \Delta d_i} - 1}{\overline{d}_i} \right) \tag{14}$$

The choice equation is a nonlinear function of risk aversion and claim rate, and there is important heterogeneity in the population in both dimensions. The equation makes clear why the deductible choice is more than just a Probit regression. Both unobserved risk aversion (υ_i) and claim rate (ε_i, through λ_i) enter the right hand side, thereby forcing us to integrate over the two-dimensional region in which the model predicts a choice of a low deductible.

A natural way to proceed is to estimate the model by Maximum Likelihood, where the likelihood of the data as a function of the parameters can be written by integrating out the latent variables, namely

$$L(claims_i, choice_i|\theta) = \Pr(claims_i, choice_i|\lambda_i, r_i) \Pr(\lambda_i, r_i|\theta) \tag{15}$$

which, imposing the distributional assumptions above, can be written as

$$\begin{gathered} L(choice_i, claims_i) = \\ = \left\{ \left(\int_0^\infty \left[\int_{r_i^*(\lambda)}^\infty f(\lambda, r|\theta) dr \right] d\lambda \right) \left(\int_0^\infty \left[\left(\frac{(\frac{\lambda}{t_i})^{claims_i} e^{-\frac{\lambda}{t_i}}}{claims_i!} \right) \int_{r_i^*(\lambda)}^\infty f(\lambda, r|\theta) dr \right] d\lambda \right) \right\}^{choice_i} \times \\ \times \left\{ \left(\int_0^\infty \left[\int_0^{r_i^*(\lambda)} f(\lambda, r|\theta) dr \right] d\lambda \right) \left(\int_0^\infty \left[\left(\frac{(\frac{\lambda}{t_i})^{claims_i} e^{-\frac{\lambda}{t_i}}}{claims_i!} \right) \int_0^{r_i^*(\lambda)} f(\lambda, r|\theta) dr \right] d\lambda \right) \right\}^{1-choice_i} \end{gathered} \tag{16}$$

where $f(\lambda, r|\theta)$ is the probability density function of the bivariate Lognormal distribution with parameters θ.

While formulating the empirical model using likelihood may help thinking about the data generating process, using Maximum Likelihood (or GMM) for estimation is computationally cumbersome. This is because in each iteration it requires evaluating a separate integral for each individual in the data. In contrast, Gibbs sampling is quite attractive in such a case. Using data augmentation of latent variables (Tanner and Wong, 1987), according to which we simulate (λ_i, r_i) and later treat those simulations as if they are part of the data, one can avoid evaluating the complex integrals by just sampling from truncated Normal distributions, which is much less computationally demanding (e.g. Devroye, 1986). This feature, combined with the idea of a "sliced sampler" (Damien et al., 1999) to sample from an unfamiliar posterior distribution, makes the use of a Gibbs sampler quite efficient for our purposes. Finally, the normality assumption implies that $\ln \lambda_i | r_i$ (and, similarly, $\ln r_i | \lambda_i$) follows a (conditional) Normal distribution, allowing us to restrict attention to univariate draws, further reducing the computational burden.

Appendix A provides all the technical details of our Gibbs sampler, including the conditional distributions and the (flat) prior distributions we use. The basic intuition is that conditional on observing (λ_i, r_i) for each individual we have a simple linear regression model with two equations. The tricky part is to generate draws for (λ_i, r_i). We do this iteratively. Conditional on λ_i, the posterior distribution for $\ln r_i$ follows a truncated Normal distribution, where the truncation point depends on the menu individual i faces, and its direction (from above or below) depends on individual i's deductible choice. This is similar to a simple Probit, except for one minor difference. In a standard Probit the level of the latent variable is typically not identified, so the variance of the normally-distributed error term is normalized to one. In contrast, in our setting the structural assumptions provide us with an alternative normalization, thus identifying the variance of the error term. The final step is to sample from the posterior distribution of λ_i conditional on r_i. This is more complicated as we have both truncation which arises from adverse selection (just as we do when sampling for r_i) as well as the claim information, which provides additional information about the posterior of λ_i. Thus, the posterior for λ_i takes an unfamiliar form. To sample from this distribution we use a "sliced sampler," a statistical method proposed by Damien et al. (1999).

We use 100,000 iterations of the Gibbs sampler. It seems to converge to the stationary distribution after about 5,000 iterations. Therefore, we drop the first 10,000 draws and use the last 90,000 draws of each variable to report our results. Note that each iteration involves generating separate draws of (λ_i, r_i) for each of the 105,800 individuals. 100,000 iterations of the whole algorithm (coded in Matlab) take about 60 hours on a Dell precision 530 workstation.

3.3 Intuition for Identification

The goal of this section is to provide intuition for the variation in the data that identifies the model, and to highlight the assumptions that are essential for identification vis-a-vis those that are only made for computational convenience (making them, in principle, testable). We discuss the identification conditional on covariates, so one can think of the discussion as being applied for a set of individuals who are identical in their observable variables (beyond deductible choices and claims). While a formal identification proof is outside the scope of this paper, we numerically verified that the estimated model is indeed (parametrically) identified by simulating data and obtaining back the pre-set parameters for various parameterizations and different initial values.

The main difficulty in identifying the model arises from the gap between the risk type (the λ_i's in our model), which is used by individuals when choosing a deductible, and the realization of risk, which is the number of claims we observe. This identification problem is similar to the one faced by Cardon and Hendel (2001). Cardon and Hendel use the variation in coverage choice (analogous to our choice of deductible) to identify the variation in health-status signals (analogous to our risk-types) from the variation in health expenditure (analogous to our number of claims). They can rely on the coverage choice to identify this because they assume a particular (i.i.d logit) form of heterogeneity in preferences. We take a different approach, as our main goal is to estimate (rather than assume) the distribution of risk preferences. We identify between the variation in risk types and in risk realizations using our distributional assumptions. This allows us to use the coverage (deductible) choice as an additional layer of information, which identifies unobserved heterogeneity in risk aversion.

Thus, the key assumption in the identification of the model is that the distribution of risk types can be uniquely backed out from the claim data alone. Any distributional assumption that satisfies this property would be sufficient to identify the distribution of risk aversion. As is customary in the analysis of count processes such as ours, we make a particular parametric assumption, assuming that claims are generated by a Lognormal mixture of Poisson distributions. Using a mixture has two advantages. First, it enables us to account for adverse selection through variation in risk types. Second, it allows the model to better fit the fatter tails of the claim distribution compared to the tails generated by a simple Poisson process.[33]

Once we make the distributional assumption that we can estimate the distribution of risk types only from claim data, the marginal distribution of risk aversion (and its relationship to the distribution of risk types) can be, in principle, nonparametrically identified. This is due to the exogenous variation in the offered menus, which is discussed in Section 2. As mentioned earlier, variation in the deductible cap over time and some experimentation with the pricing policy provide variation in the menu faced by two identical (on observables) individuals, who purchased insurance from the company at different times. Different menus of deductible-premium options lead to different indifference sets (similar to the one depicted in Figure 3), which often cross each other and nonparametrically identify the distribution of risk aversion and the correlation structure, at least within the region in which the indifference sets vary. At the tails of the distribution, as is typically the case, there is no data, so we have to rely on parametric assumptions or to use bounds.

[33] An alternative is a Negative Binomial distribution, which generalizes Poisson to allow for overdispersion and is often used to model count processes. In general, it will allow more overdispresion to be explained by the distribution rather than by heterogeneity, thereby giving less room for possible adverse selection. This is likely to increase the estimated heterogeneity (and therefore the mean) of estimated risk aversion compared to our benchmark model.

The assumption of Lognormality we use throughout the paper is only made for computational convenience.

Let us now provide simple intuition for the (parametric) identification mechanism. To keep the intuition simple, let us take the bivariate Lognormal distribution as given and, contrary to the data, assume that all identical individuals face an identical menu of deductible-premium combinations. Suppose also that the maximum number of claims observed for each individual is two, and that all individuals are observed for exactly one year.[34] In such a case, the data can be summarized by five numbers. Let α_0, α_1, and $\alpha_2 = 1 - \alpha_1 - \alpha_0$ be the fraction of individuals with zero, one, and two claims, respectively. Let φ_0, φ_1, and φ_2 be the proportion of individuals who chose low deductible within each "claim group." Given our distributional assumption about the data generating process of the claim distribution, we can use α_0 and α_1 to uniquely identify the mean and variance of the Lognormal distribution of risk types. Given this distribution, we can (implicitly) construct a posterior distribution of risk types for each claim group, namely $F(\lambda|r, claims = c)$, and integrate over it when predicting the deductible choice. This will provide us with three additional moments, each of the form

$$E(\varphi_c) = \int\int \Pr(choice = 1|r, \lambda)dF(\lambda|r, claims = c)dF(r) \tag{17}$$

for $c = 0, 1, 2$. These moments will then uniquely identify the three remaining parameters of the model, namely the mean and variance of the risk aversion distribution, as well as the correlation coefficient.

Let us finally provide more economic content to the above identification argument. Following the same example and conditional on the (already identified) distribution of risk types and the corresponding posteriors, one can think about the deductible choice data as providing a line described by the various φ_c's. The absolute level of the line identifies the average level of risk aversion. In the absence of correlation between risk and risk aversion, the slope of the line identifies the variance in risk aversion. With no correlation, the slope should always be positive (due to adverse selection), but the line will be flatter as the variance of risk aversion is higher, as more of the deductible choice will be attributed to variation in risk aversion, which is uncorrelated with claims. Finally, the correlation parameter can be thought of as identified by the curvature of the line. The more convex (concave) the line is, the more positive (negative) the correlation parameter. For example, if $\varphi_0 = 0.5$, $\varphi_1 = 0.51$, and $\varphi_2 = 0.99$ it is very likely that the variance of risk aversion is high (explaining why φ_0 and φ_1 are so close) and the correlation between risk and risk aversion is highly positive (explaining why φ_2 is not also close to φ_1). In contrast, if $\varphi_0 > \varphi_1$ it must mean that the correlation between risk and risk aversion is negative, which is the only way the original positive correlation induced by adverse selection can be offset. This intuition also clarifies that identification relies on observations with multiple claims (or different policy durations).

[34] In the data we have, of course, more degrees of freedom. We observe up to five claims per individual, we observe continuous variation in the policy durations, and we exploit distributional restrictions across individual with different observables.

To summarize, one should note that the extent of the positive (or negative) correlation is strongly related to the structural model for deductible choice described earlier. The data (see Table 2B) provide direct correlation between deductible choice and risk (claims). The structural assumptions allow us to explain how much of this correlation can be attributed to adverse selection. The remaining correlation is therefore attributed to correlation in the underlying distribution of risk and risk aversion.

4 Results

4.1 Descriptive Analysis

Cohen (2005) provides "reduced form" evidence for the existence of adverse selection in our data using a version of the bivariate Probit test suggested by Chiappori and Salanie (2000). Table 3 and Table 4A repeat some of these regressions and provide some reduced-form analysis of the relationship between the observables and our two left-hand side variables, the number of claims and the deductible choice. Table 3 reports the estimates from a Poisson regression of the number of claims on the covariates. This regression is closely related to the risk equation we estimate in our benchmark model. It shows that older people, women, and people with academic education are less likely to have an accident. Bigger, more expensive, older, and non-commercial cars are more likely to be involved in an accident. Driving experience and variables associated with less intense use of the car reduce accident rates. As could be imagined, claim propensity is highly correlated over time: past claims are strong predictor of future claims. Young drivers are about 50% more likely to be involved in an accident, with young men significantly more than young women. Finally, as indicated by the trend in the estimated year dummies, the accident rate significantly declined over time. Part of this decline is likely due to the decline in accident rates in Israel in general (in particular, traffic fatalities and counts of traffic accidents in Israel fell by 11% and 18% during 1998 and 1999, respectively). This decline might also be partly due to the better selection of individuals the company obtained over time, as it gained more experience; over time, the company might have improved its ability to identify and price out the more risky potential customers.

Table 4 presents estimates from simple Probit regressions in which the dependent variable is equal to 1 if the policy holder chose a low deductible, and is equal to zero otherwise. In general, the coefficients should proxy for risk attitudes. More risk averse individuals should be more likely to choose low deductibles. This is not precise, however, as the price of risk varied with demographics. Thus, it may be that a certain coefficient is positive not because of its association with higher risk aversion, but because it is associated with risk, which is under-priced by the company. Other columns of Table 4A control for risk and prices, and some of the coefficients indeed change. Ultimately, these interpretation problems are the reason one needs a more structural model, such as the one we estimate below. With this qualification in mind, Table 4A suggests that older people are less risk averse, while women, owners of expensive cars, and individuals who had recent claims are more risk averse. In this regression we again observe a strong trend over time. Fewer and fewer policy holders chose the low deductible as time went by. One reason for this trend, according to the company executives, is that over time the company's sales persons were directed to mainly focus on the "default," regular deductible-premium option.[35]

Table 4C presents a structural interpretation of the simple Probit regression. This can be thought of as a restriction of the benchmark model, which does not allow unobserved heterogeneity in risk exposure. In such a case consumers have no private information about their risk type, and the risk type can thus be estimated directly from the data. The structure of the model dictates the functional form in which the predicted risk and the deductible-premium combinations enter into the deductible choice. Under the Lognormality assumption for the risk aversion distribution, this just means that the additional variable is $log(\frac{\Delta p_i/(\widehat{\lambda}(X_i)\Delta d_i)-1}{\overline{d}_i})$. The structural assumptions imply that the coefficient on this variable is -1, thus freeing up the normalization of the Probit error term. While the signs of the estimated coefficients are similar in Table 4C and in the benchmark model presented below, the restricted version of the model suggests much higher effects, and much higher significance levels, for all coefficients. It also suggests a significantly higher unobserved heterogeneity in risk aversion. It is clear that the full estimation of the benchmark model rejects this restriction on the model.

Finally, to get a sense of the levels of absolute risk aversion implied by the data, one could use a simple back-of-the-envelope exercise. We compute unconditional averages of Δp, Δd, λ, d^h, d^l, and $\overline{d}$ (see Table 2A),[36] and substitute these values in equation (7). The implied coefficient of absolute risk aversion from this exercise is $2.9 * 10^{-4}\ NIS^{-1}$. One can also implicitly solve for the coefficient of risk aversion using the CARA specification in equation (8), which gives a slightly lower value of $2.5 * 10^{-4}$. Ignoring nonlinearities, we can go on and think of this level as the average cutoff point, implying that about 18% of our policy holders have a coefficient of absolute risk aversion exceeding it. To convert to US dollar amounts, one needs to multiply these figures by the average exchange rate (3.52), resulting in an average indifference point of $1.02 * 10^{-3}$ ($8.8 * 10^{-4}$ in the CARA case). This figure can be compared to two other similar figures reported in the literature. Metrick (1995) imposes the CARA utility function, and estimates the coefficient of absolute risk aversion (for a representative player in "Jeopardy!") to be $6.6 * 10^{-5}$, which is 13 to 15 times lower than the figures above. Gertner (1993) finds a lower bound of the CARA coefficient (for a representative player in "Card Sharks") to be $3.1 * 10^{-4}$, which is 3-4 times lower than the figures above. One can continue with back of the envelope exercises, and multiply our figure of absolute risk aversion by the average disposable annual income, which was about 50,000 NIS at the time, to obtain a measure for the coefficient of relative risk aversion. This implies that the 82nd percentile in the distribution of the coefficient of relative risk aversion is about 13-15. This, of course, also ignores the fact that

[35] One should note that such biased marketing efforts will bias consumers against choosing the low deductible, thus making them look less risk averse. This would make our estimate a lower bound on the true level of risk aversion. If only sophisticated consumers could see beyond the marketing effort, and this sophistication is related to observables (e.g. education), the coefficients on such observables would be biased upwards. Such biases can be evaluated by running the model on each year separately.

[36] The unconditional λ is computed by maximum likelihood, using the data on claims and observed durations of the policies.

absolute risk aversion and wealth are likely to be correlated. All these exercises are, in general, rather crude and imprecise, and do not account for nonlinearities, heterogeneity, and endogeneity. This is exactly why we need the more structural model we develop is this paper.

4.2 Estimation Results

Risk Aversion and Individual Characteristics Table 5 presents the results from the benchmark model. The risk aversion coefficients in the two right columns of Table 5 are, in our view, one of the main contributions of the paper. They show how the level of absolute risk aversion is related to the demographic characteristics of individuals. As the dependent variable is in natural logarithm, coefficients on dummy variables can be directly interpreted as approximate percentage changes. Table 6 repeats the same exercise for a CARA specification.

Our results indicate that women are more risk averse than men. In particular, women have a coefficient of absolute risk aversion about 16% greater than that of men. These results are consistent with those of Donkers et al. (2001) and Hartog et al. (2002). The effect of age and marital status is not significant.[37] One exception is divorced individuals who appear to be less risk averse, which seems reasonable. A somewhat surprising result of our analysis is that variables which are likely to be correlated with income or wealth seem to have a positive coefficient, indicating that wealthier people have a higher level of absolute risk aversion. This is true for individuals with post high school education, as well as for owners of more expensive cars. In unreported regressions, we also find that the elasticity of absolute risk aversion with respect to gross monthly income, as measured by average income among households living in the same zip code, is positive 0.35, and is highly significant. At first glance – but only at first glance – these results may appear to be inconsistent with the widely held belief that absolute risk aversion declines with wealth. It is important, however, to distinguish between two questions: (i) whether, for a given individual, the vNM utility function exhibits decreasing absolute risk aversion; and (ii) how risk preferences vary across individuals. Our results do not at all speak to the first question and should not be thought of as a test of the decreasing absolute risk aversion property. This property can only be tested by observing the same individual making multiple choices at different wealth levels. Rather, our results are more closely related to the second question, i.e. to a comparison among individuals. The results indicate that individuals with greater wealth have utility functions that involve a greater degree of risk aversion. It might be that risk aversion, or individual characteristics that are correlated with it, lead individuals to save more, to obtain more education, or to take other actions that lead to greater wealth.

One may be tempted to interpret the positive wealth effects we find above as an indirect indication for credit constraints. Wealthier individuals are less credit constrained, and therefore can afford to purchase more insurance. We do not share this interpretation for two reasons. First, this is an unconditional interpretation. Note, however, that the insurance company observes these proxies for wealth and conditions on them when setting prices. Since the willingness to pay for

[37] While age has significant effects in both Donkers et al. (2001) and Hartog et al. (2002), it takes different signs.

insurance is likely to be correlated with the willingness to pay for the additional insurance provided by the low deductible option, premia already reflect this variation. We condition on the level of the premium, and therefore the wealth coefficients we find are conditional on premia, rather than unconditional. Second, one should note that paying less ex-ante implies paying more ex-post, so applying the credit constraint argument only for the ex-ante payment but not for the probabilistic ex-post deductible payments has no theoretical foundation. Essentially, the setup of the model links the ex-ante decisions with the ex-post losses, which are both driven by the curvature of the vNM utility function. This is exactly how we interpret our findings.

Let us make several additional observations. First, while the owners of more expensive cars appear to have both higher risk exposure and higher levels of risk aversion, owners of bigger cars have higher risk exposure but lower levels of risk aversion. This should indicate that the structure of the model itself does not necessarily constrain the relationship between the coefficients in the two equations. Rather, it is the data that speak up. Second, it is interesting to note that individuals who are classified by the insurer as "good drivers" indeed have lower risk, but also appear to have lower risk aversion. This result is somewhat similar to the positive correlation between unobserved risk and unobserved risk aversion, which we report below. We discuss its interpretation later. Third, the results suggest that policyholders who tend to use the car for business are less risk averse. This finding might be due to the fact that the uninsured costs of accidents occurring to such policyholders might be borne by their employer or might be tax deductible. We find that policyholders who reported three full years of past claim history are more risk averse, but are not different in their risk exposure. The attitude that leads such policyholders to comply with the request to (voluntarily) report full three years of claim history is apparently, and not surprisingly, correlated with higher levels of risk aversion. In contrast, while past claims indicate high risk, they have no significant relationship with risk aversion. Finally, one should note the strong trend towards lower levels of risk aversion over time. This is a replication of the Probit results reported and discussed at the end of the preceding section.

The Level of Risk Aversion Our results enable us to estimate the levels of absolute risk aversion in the population we study. Since we use Gibbs sampler and augment the latent coefficients of absolute risk aversion, we can directly obtain the posterior distribution of various moments of the distribution. Note that these estimates cannot be read directly from the regression results in Table 5. Since the dependent variable is in natural logarithm, one needs to account for the distribution of the observables when computing moments of the distribution of r_i. In Table 7 we report the point estimates of the (unconditional) mean and median from this distribution. The implied risk aversion of the mean individual is 0.0016, which is four times greater than the back-of-the-envelope calculation presented in the end of the previous section. As we assume a Lognormal distribution and estimate a relatively high coefficient of σ_r, the estimates also imply significantly lower estimates for the median level of risk aversion.

In Table 7 we also present two ways to interpret the estimates, as well as comparisons to a CARA specification, to an incomplete-information specification (see later), and to other comparable figures in the literature (Gertner, 1993; Metrick, 1995; Holt and Laury, 2002). The estimate suggests that an average quadratic utility maximizer[38] will be indifferent about participating in a lottery in which she gains 100 dollars with probability 0.5 and loses 61.3 dollars with probability 0.5. A CARA specification suggests much lower value for average risk aversion, but one that is still higher than other results in the literature. The reason that a CARA specification affects the levels so much is because of its relatively high third derivative. This specification introduces an additional (precautionary) incentive to choose low deductibles, thus not requiring very high levels of the coefficient of absolute risk aversion to explain the observed deductible choices. Finally, Table 7 shows that the incomplete information model does not change the main conclusions.

Let us briefly discuss the relevance of the comparison to Gertner (1993) and Metrick (1995). There are two ways in which one can reconcile the differences between the estimates. First, as already discussed, both of these papers measure risk aversion for television show participants; these are highly selected groups in a rather "risk-friendly" environment.[39] Second, the magnitudes of the stakes are higher. Their television show participants make bets over at least several thousands of dollars, while our average individual risks much lower stakes, at the range of one hundred dollars. Thus, the difference in the results may be due to the issues raised in Rabin (2000) regarding the comparability of behavior across different contexts and bet sizes. People may behave differently, i.e. exhibit different levels of absolute risk aversion, for different sizes of bets. In fact, in a similar fashion to Rabin's (2000) exercise, if we apply our estimates for fifty-fifty bets of much bigger size, the implied certainty equivalence is extremely low. As the goal in this paper is to estimate risk attitudes of insurees in the context of the decisions they have to make, we do not pursue this extrapolation exercise any further, and feel comfortable to report them as consistent estimates for bets at the one hundred dollar range.

A different way to quantify our estimate is by reporting it in relative terms. Following the literature (e.g. Gertner, 1993), we do so by multiplying the estimated coefficient of absolute risk aversion by the average annual income in Israel during the observation period. Under the assumption that annual income is a good proxy for the relevant wealth at the time of decision making, this product would be a proxy for the coefficient of relative risk aversion. As discussed in the introduction, there are many good reasons to question such an exercise on structural grounds. We provide it mainly as a way to compare our estimates to those found in the literature. As Table 7 indicates, our benchmark specification results in an implied coefficient of relative risk aversion of about 82. A CARA specification results in a lower coefficient of 9.8. Both of these figures, however, are significantly higher than the widely used estimate of a low single-digit coefficient of relative risk aversion.

Finally, we should note that, as reported in Table 7, our estimates for the average individual are much higher than those for the median individual. Although the mean is always greater than the

[38] For such an individual the second-order Taylor expansion we use in Section 3.1 is exact.

[39] We suspect that individuals who participate in television game shows are more adventuresome than the general population. Moreover, knowing that the audience might wish to see them keep betting is likely to further encourage participants to take risks.

median under the Lognormal distribution, the big difference we find is not imposed. In principle, we could have obtained a high level of risk aversion with less heterogeneity, thereby leading to a smaller difference between the mean and the median (the estimated distribution of risk types is an example). The highly skewed estimated distribution of risk aversion may reflect the fact that, indeed, most people are almost risk neutral with respect to bets of these relatively small sizes, but a small fraction of individuals is extremely risk averse.

The Risk Regression The risk coefficients in the first two columns of Table 5 (and, similarly, Table 6) provide information on the relationship between observables and exposure to risk. They indicate that the likelihood of an accident is smaller for people with academic education. Bigger, more expensive, and non-commercial cars are more likely to be involved in an accident. Driving experience reduces accident rates, as do measures of less intense use of the car. Claim propensity is highly correlated over time. The voluntary report of past claims is a strong predictor of future claims. Young drivers are more likely to be involved in an accident, with young men significantly more so than young women.

It is worth noting that the risk regression in Table 5 produces results that are similar to those of the simpler Poisson regression reported in Table 3. Although some of the coefficients lose significance, the magnitude of most coefficients is quite similar to those presented in Table 3. The similarity between these two sets of results is to be expected, as there is very little new information that the structural model incorporates into the risk regression. As discussed in the previous section, the risk regression is identified only from the data on claims, so incorporating the information on deductible choice does not qualitatively change the conceptual identification strategy. If the results were not similar, this would have been an indication for a misspecification of the model. The slight differences between the risk regressions in Table 3 and Table 5 are mainly driven by the structural assumptions. First, the benchmark model estimates a Normal mixture of Poisson models, rather than a single Poisson model. By incorporating the fatter tails of the claim distribution, it slightly changes the results, increases the standard errors, and decreases the average predicted claim rate. Second, the information on deductible choice slightly helps us in obtaining more precise estimates through the correlation structure between the error terms in the two equations.

The Relationship between Unobserved Risk and Unobserved Risk Aversion Table 5 allows us to make observations about the relationship between risk and risk aversion. First, our results enable us to assess the relative importance of unobserved heterogeneity of both dimensions. In the population we study, the unobserved heterogeneity in risk aversion (σ_r) is much greater than the unobserved heterogeneity in risk (σ_λ). This is true both in absolute terms and after normalizing by the corresponding mean level. This may indicate that selection on risk aversion is more important in our data than selection on risk, i.e. adverse selection. One should note, however, that the right metric to use for such statements is not entirely clear, as one should project these estimated variances onto the same scale of, say, willingness to pay or profits. Therefore, we relegate the discussion of this issue to Section 4.3.

Furthermore, Table 5 also indicates a strong and significant positive correlation of 0.86 between unobserved risk aversion and unobserved risk. This result might be viewed as surprising because it is natural to think that risk aversion with respect to financial decisions is likely to be associated with a greater tendency to take precautions, and therefore with lower risk. Indeed, a recent paper by Finkelstein and McGarry (2003) supports such intuition by documenting a negative correlation between risk aversion and risk in the market for long term care insurance.[40] Our market, however, might be special in ways that could produce a positive correlation. First, in contrast to most insurance markets where a poliycholder's risk depends on the policyholder's precautions but not on the precautions of others, accident risk in the auto insurance market is a result of an interaction between one's driving habits and those of other drivers. In the auto insurance market, driving too slow or too carefully may actually expose a policyholder to a greater risk. An indication that something like this may be going on is the negative coefficient on the "good driver" variable in the risk aversion regression of Table 5.

Second, the correlation coefficient may be highly sensitive to the particular way we measure risk and risk aversion. There are many unobserved omitted factors that are likely to be related to both dimensions. For example, the extent to which individuals drive more carefully may not be the primary determinant of the risk posed by an individual policyholder. The intensity of vehicle use, for example, might be a more important determinant of risk. If individuals who are more risk averse also drive more miles per year, a positive correlation between risk and risk aversion could emerge. Thus, our results caution against assuming that risk and risk aversion are always negatively correlated. Whether this is the case may depend on the characteristics of the particular market one studies, and on the particular measure for risk. Indeed, one can use estimated annual mileage to control for one omitted variable that may potentially work to produce a positive correlation between risk aversion and risk. Despite its partial coverage in the data and being considered (by the company) as unreliable,[41] controlling for annual mileage reported by policyholders reduces the estimated correlation to about 0.7. We view this result as consistent with the possibility that underlying unobserved factors that affect risk play an important role in generating the estimated positive correlation between risk and risk aversion.

Finally, one should note that while the correlation parameter we estimate is extremely high, the implied *unconditional* correlation between risk and risk aversion is much lower, and is less than 0.2 across all reported specifications. This is because the coefficients on the same regressor (for example, the size of the car or whether the car is used for business) often affect risk and risk aversion in opposite directions.

Additional Specifications We have tried several other specifications in order to check the robustness of the results to different assumptions. First, we estimate the model with a CARA specifi-

[40] See also Israel (2005).

[41] Insurance companies typically do not use this self-reported mileage estimates as they are considered unreliable. While companies could verify these estimates at the time of a claim, such reports are hard to enforce. An individual can always claim that her ex-ante estimate was lower than it turned out to be.

cation. The results are reported in Table 6. Due to its high third derivative and flatter indifference set, the CARA assumption results in lower estimates of the average level of risk aversion and the variance of unobserved heterogeneity. The rest of the results are qualitatively very similar. Most of the covariates take similar coefficients in signs, magnitudes, and statistical significance. The risk regression changes very little, and the correlation remains virtually unchanged.

To account for potential heterogeneity in the information individuals have with respect to their risk level, we try two alternatives. First, we estimate the model for experienced drivers only, namely for drivers with ten years or more of driving experience. The underlying assumption is that such experienced drivers, unlike new drivers, know their own risk rate much better, and our model, which assumes that an individual perfectly knows her risk type, applies better for experienced drivers. The results (Table 8) are virtually unchanged, and the correlation coefficient is only slightly lower (0.78). Second, we estimate a different specification of the model, which structurally allows for incomplete information of individuals with respect to their risk types. We do so by assuming that individuals are Bayesian and update their risk types over time, given information about the number of claims they make each year. While we do not observe claim histories of individuals, we can simulate such a history and integrate over the simulation draws. Thus, individuals' information would be related to their true type, and would be more precise with longer driving histories. Overall, this seems to us an extreme version of incomplete information, as there are many other sources through which individuals can learn about their own types, and thereby have better information about their types than what we estimate them to have. The end of Appendix A provides more details of this specification and Table 9 reports the results. The qualitative results are very similar.

As we already mentioned, we also tried estimating the model when controlling for (self-reported) estimated mileage. Due to partial coverage, this specification results in omitting almost 50% of the observations, which may introduce a selection bias. Nevertheless, the results are quite similar, with the correlation coefficient going down to about 0.7. The coefficient on mileage is significant but small; this result implies that the elasticity of claim rate with respect to mileage is about 10%. We interpret this finding as consistent with the possibility that self-estimated mileage is not a particularly reliable variable, meaning that the low estimated coefficient on it is due to "errors in variables," biasing it towards zero. One could speculate that more precise mileage data would have led to a further reduction in the correlation coefficient, which is consistent with our earlier discussion.

As already mentioned, we estimated the model incorporating average income in the same zip code.[42] While income obtains a positive and significant coefficient in the risk aversion equation (and an insignificant coefficient in the risk regression), it does not much affect the reported results. The correlation coefficient is about 0.82, the level of risk aversion remains fairly similar and unobserved heterogeneity is slightly higher. We also tried to estimate the model separately for each of the five years. For the last two years of data, we encountered convergence problems. This may be due to insufficient exogenous variation in the data (see Figure 1), which leads to weak identification. For

[42] The reason we do not use this income variable in the benchmark specification is its imperfect coverage. It requires us to omit almost 20% of the individuals.

the first three years most of the estimated coefficients are qualitatively stable over time, although their magnitude does tend to vary, and many of them lose significance. The average level of risk aversion increases over time, and so does its estimated variation. The correlation coefficient, which may be of particular interest, is consistently positive and significant, although it takes lower values than in the benchmark (pooled) regression; its values are 0.25 (0.12), 0.43 (0.11), and 0.58 (0.08) for the first, second, and third year, respectively (standard deviations in parentheses). It is important to note that variation in the coefficients over time is to be expected, due to selection. It is likely that different types of individuals have approached the company as it gained more experience. We view the results from the benchmark model as reflecting an average customer of the company over its first five years of operation.

4.3 Counterfactuals

We focus in this section on the analysis of a profit-maximizing choice of insurance contracts in the presence of the two dimensions of private information. As our results only represent the distribution of risk and risk aversion in the population of customers we observe, and as we have no information about the outside option of these customers, we hold this distribution fixed throughout our counterfactual exercises. To make the counterfactual exercise still meaningful, it is necessary to minimize the potential for a significant inflow or outflow from the observed population, as we change the menu of deductible-premium combinations.

In light of this concern, we proceed by making the simplifying assumption that individuals make their choices sequentially. They first choose the insurance provider by only observing the "regular" deductible-premium combination they are offered. Once they decide to buy a policy from the insurer, they decide which deductible-premium combination to purchase. This is clearly a strong assumption, but, in our view, it is a reasonable approximation of reality, as the "regular" deductible is the one always advertised and initially quoted, while the other options are only revealed once the potential customer and the insurance sales person get into details. As a consequence of this assumption, we do not analyze the optimality of the regular premium and regular deductible levels. These are assumed to be dictated by competitive conditions. Instead, we focus our analysis on the optimality of the choice of the level and price of the low deductible option. One should also keep in mind that this assumption should hold much better locally than globally. As we investigate deductible-premium combinations that are further away from those that are actually offered, failure of the assumption may become more prevalent.

Consider a particular individual. As far as the company is concerned, this individual can be represented by a random draw of (λ_i, r_i) from the conditional distribution of risk and risk aversion:

$$\begin{pmatrix} \log(\lambda_i) \\ \log(r_i) \end{pmatrix} \sim N\left(\begin{pmatrix} \overline{\log(\lambda)} \\ \overline{\log(r)} \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{pmatrix} \right) \tag{18}$$

where all parameters should be thought of as conditional on observables. When analyzing the optimal menu to offer such an individual, the company is assumed to be risk neutral and to maximize expected profits. Below we analyze how the price and level of the low deductible option offered by the insurance company affect the company's expected profits.

Suppose the company only offered the "regular" deductible-premium combination, (d_h, p_h). Let the expected profits from this strategy be π_0. Consider now the profits of the firm from offering a "low" deductible-premium combination, (d_l, p_l) with $d_l < d_h$ and $p_l > p_h$. We will analyze the optimality of the decision (d_l, p_l). As will become clear soon, it is easy to use a change in variables and analyze the choice of $\Delta d = d_h - d_l$ and $\Delta p = p_l - p_h$. Expected profits are now given by:

$$\max_{\Delta d, \Delta p} \{\pi_0 + \Pr(r_i > r_i^*(\lambda_i; \Delta d, \Delta p)) \left[\Delta p - \Delta d \cdot E(\lambda_i | r_i > r_i^*(\lambda_i; \Delta d, \Delta p)\right]\} \tag{19}$$

The trade-off in the company's decision is straightforward. Each new customer who chooses the low combination pays an additional Δp up-front, but saves Δd for each accident she is involved in. This translates into two effects that enter the company's decision problem. The first is similar to a standard pricing problem: higher (lower) price difference (deductible difference), Δp (Δd), leads to a higher markup (on those individuals who select the "low" combination), but to lower quantity (or probability of purchase) as fewer individuals elect to choose the low deductible. This effect enters the profit function through $\Pr(r_i > r_i^*(\lambda_i; \Delta d, \Delta p)) \equiv D(\Delta d, \Delta p)$. The second, composition effect arises because of adverse selection. As the price of the low deductible increases, those individuals who still elect to choose the low combination are, *ceteris paribus*, those with higher risk. This effect enters through $E(\lambda_i | r_i > r_i^*(\lambda_i; \Delta d, \Delta p)) \equiv \overline{\lambda}(\Delta d, \Delta p)$. Its magnitude and sign depend on the relative heterogeneity of λ_i and r_i and on the correlation between them.

First order conditions are given by

$$0 = \frac{\partial D(\Delta d, \Delta p)}{\partial \Delta p}\left[\Delta p - \Delta d \cdot \overline{\lambda}(\Delta d, \Delta p)\right] + D(\Delta d, \Delta p)\left[1 - \Delta d \cdot \frac{\partial \overline{\lambda}(\Delta d, \Delta p)}{\partial \Delta p}\right] \tag{20}$$

$$0 = \frac{\partial D(\Delta d, \Delta p)}{\partial \Delta d}\left[\Delta p - \Delta d \cdot \overline{\lambda}(\Delta d, \Delta p)\right] - D(\Delta d, \Delta p)\left[\overline{\lambda}(\Delta d, \Delta p) + \Delta d \cdot \frac{\partial \overline{\lambda}(\Delta d, \Delta p)}{\partial \Delta d}\right] \tag{21}$$

Since neither $D(\Delta d, \Delta p)$ nor $\overline{\lambda}(\Delta d, \Delta p)$ have a closed-form solution, we will analyze this decision problem graphically, where $D(\Delta d, \Delta p)$ and $\overline{\lambda}(\Delta d, \Delta p)$ are numerically computed using simulations from their joint distribution.[43]

We will illustrate our analysis by using the mean individual in the data. Such an individual is faced with a regular combination of $(p_h, d_h) = (3190, 1595)$ New Israeli Shekels (NIS). The current low combination offered to her is $(p_l, d_l) = (3381, 957)$ NIS, i.e. $(\Delta p, \Delta d) = (191, 638)$. According to our benchmark estimates (Table 5), this individual is viewed by the company as a random draw from

$$\begin{pmatrix} \log(\lambda_i) \\ \log(r_i) \end{pmatrix} \sim N\left(\begin{pmatrix} -1.57 \\ -11.63 \end{pmatrix}, \begin{pmatrix} (0.172)^2 & 0.861 \cdot 0.172 \cdot 2.986 \\ 0.861 \cdot 0.172 \cdot 2.986 & (2.986)^2 \end{pmatrix}\right) \tag{22}$$

In each figure discussed below, we also present similar exercises for cases of zero correlation and negative correlation (opposite sign, same magnitude of 0.861) between risk and risk aversion. This should help in the interpretation of the various forces in play.

[43] These first order conditions could, in principle, be used as "supply side" moment conditions for estimation.

To get intuition for the different effects, Figure 4 presents the estimated distribution in the space of (λ_i, r_i). A small increase (decrease) in Δp (Δd) shifts the indifference set up and to the right, thereby making some marginal individuals, who were previously just to the right of it, switch to choosing the regular deductible. The demand trade-off is just the comparison between the marginal loss of the company from all the marginal individuals who no longer buy higher coverage vis-a-vis the higher profits made from the infra-marginal individuals who still elect to choose higher coverage. Figure 4 also helps in illustrating the effect of adverse selection and the importance of the correlation coefficient. As the menu shifts to the right, the positive correlation implies that the marginal individuals have higher risk than the average. This means that "losing" them (namely, having them buy less coverage) is not as costly for the insurance company, as such individuals are, on average, more adversely selected. A negative correlation, for example, would have made these marginal individuals more valuable, thereby decreasing the incentive to increase prices or deductibles from the current levels.

Figure 5 presents the implications of the model as we vary the low deductible level, keeping the premium charged for it fixed at the true price of $\Delta p = 191$ NIS. The upper panel shows the effect on profits. It implies that the current low deductible benefit of 640 NIS results in additional annual profits of about 3.4 NIS per customer. This is about 0.34 percent of total operating profits per customer, which are about $1,000$ NIS. Note, however, that after subtracting the administrative and claim-handling costs associated with each customer and claim (which, by assumption, are independent of the deductible choice), the relative magnitude of this effect will be much higher. Note, also, that the estimates imply that the current low deductible level is suboptimal. By setting a smaller low deductible benefit of $\Delta d = 350$ NIS (which implies a 290 increase in the level of the low deductible), *additional* profits can be increased to 6.3 NIS. There is no apparent reason, of course, to limit the choice of the company to only one additional deductible level. More degrees of freedom in choosing the menu offered will lead, of course, to higher profits. In that sense, the estimates provided can be thought of as lower bounds.

The other two panels of Figure 5 present the way the effect on profits is generated, by analyzing the effect of the deductible level on the demand for low deductible, $D(\Delta d, \Delta p)$, and on the composition effect, $\overline{\lambda}(\Delta d, \Delta p)$. The former is simply generated by the distribution of certainty equivalents implied by the joint distribution of λ_i and r_i (see also Landsberger and Meilijson, 1999). It has an S shape due to the joint Lognormality assumptions. The shape of the composition effect is driven by the relative variance of λ_i and r_i and by the correlation coefficient, as already discussed. As the estimates imply that most of the variation in certainty equivalents is driven by variation in r_i, the strong positive correlation implies that the composition effect is monotonically decreasing in the deductible level. As the low deductible option becomes more favorable, more people choose it, with the most risky individuals being the first.

It is interesting to see that the effect of the deductible level on the composition effect is dramatically different when the correlation between risk and risk aversion is zero or negative. With zero correlation, the two extremes of the deductible range are roughly the same, as they are solely driven by the risk aversion distribution, which is uncorrelated with risk. Only at interim levels of deductibles do we see the effect of adverse selection. Note also that in such a case, the optimal deductible benefit is higher, and at the optimum there is little difference between the risk of individuals who choose low deductibles and the rest of the population.[44] Finally, one can see that when risk and risk aversion are negatively correlated the observed relationship between the deductible level and the composition effect is mostly reversed. This is because the effect of risk aversion dominates that of adverse selection due to its higher variance. Along these lines, we have also computed the optimal pricing by the company when each of the dimensions of heterogeneity is shut down. The results are consistent with the observation that adverse selection is less important than selection on risk aversion. By ignoring adverse selection, the optimal pricing does not change much. By ignoring heterogeneity in risk aversion, the optimal pricing and the shape of the profit function is significantly different.

5 Caveats: Discussion of Unmodeled Elements

The empirical model we estimated is, of course, very stylized. Below we discuss how relaxing various modeling assumptions may affect the results. Before we do so, let us make two comments that apply to most of this section. First, each discussion below opens up an additional dimension, which we abstract from in the estimated model. Since the data, in principle, only includes important variation in two dimensions (claims and coverage choices) and since we already try to identify two dimensions of unobservables (risk and risk aversion), any additional dimension can only be identified using either richer data or additional parametric restrictions. In other words, with the existing data, any pattern that can be generated by any of the mechanisms discussed below, could be generated by our original model with sufficiently flexible distributional assumptions. This is the main reason why we prefer to discuss the effects of these additional dimensions here rather than to incorporate them in the estimated model. Where appropriate, we discuss what kind of data would have been useful to address each point.

The second comment is that we have reported two different sets of results, one regarding the parameter estimates for the distribution of risk aversion and the other regarding the couterfactual pricing exercise. As we discuss below, some assumptions may be important for one of these sets of results but not for the other. This is because the estimates of risk aversion are only important to the extent that they can be extrapolated to other decision contexts. Therefore, any idiosyncratic effect of the particular choice under consideration may make this extrapolation less accurate. In contrast, the counterfactual exercise investigates the effect of pricing within the same context, making it less sensitive to the exact interpretation of risk aversion, but potentially more sensitive to the (unobserved) outside option.

[44] This may point to another potential reason that empirical papers do not find evidence for adverse selection. Not only is adverse selection dominated by higher unobserved variation in risk preferences, it may be also mitigated by the optimal decisions of insurance companies.

Moral hazard Throughout our analysis we have abstracted from moral hazard, i.e. we assumed that λ_i can vary across individuals but is invariant to the coverage choice. There are two types of moral hazard that may play a role in this context, making λ_i higher for higher coverage (low deductible). First, individuals with less coverage, who face higher loss in the event of an accident, may take greater precaution and drive more carefully, thereby reducing their accident risk rate. Second, conditional on a claim event, people with higher deductibles are less likely to file a claim: there exists a range of claims for which filing is profitable only under a low deductible. This second effect is often called ex-post moral hazard. We discuss each effect in turn.

It seems reasonable to conjecture that, *ceteris paribus*, insured individuals will drive less carefully than uninsured ones. It may also seem reasonable that the existence of a deductible may make individuals more careful about small damages to their car (which is, in fact, one of the primary reasons for the existence of deductibles). When the coverage choice, however, always includes a deductible, and different deductibles are similar in their magnitudes, it seems less likely that driving/care behavior will be affected.[45] If driving behavior is affected by the deductible choice, this will likely bias our estimates of risk aversion downwards. To see this, note that adjusting behavior will help individuals to self insure against uninsured costs. This will make higher coverage (low deductible) less attractive, requiring individuals to be even more risk averse than we estimate them to be in order to buy higher coverage. Finally, to separately identify moral hazard would require another dimension of the data such as a panel structure, over which risk types remain fixed but coverage choices exogenously vary (see also Chiappori and Heckman, 2000).

We abstract from the second potential effect, that of ex-post moral hazard, based on our data. Data on the claim amounts show that about 99% of the claims filed by policyholders with low deductible policies were for amounts greater than the higher deductible level. In other words, if individuals filed a claim for any loss which exceeds their deductible, the above analysis suggests that 99% of the claims would have been filed under either deductible choice. Figure 6 provides more details. The above exercise may be somewhat misleading as one may be worried that when filing a claim, an individual will take into account the dynamic costs as well. The dynamic costs of filing a claim come into play through its effect on experience rating, which increases future insurance premia. These dynamic effects do not depend on the deductible level at the time of the claim, so they simply enter in an additive way.

Using our data on individuals who renew their policies with the company (these renewals are not used for estimation), we can assess how big the dynamics effects are. These data show that the price effect of a claim lasts for three years, and is highest when an individual files her second claim within a year. In such a case, she would face about 20 percent increase in her insurance premium in the subsequent year, 10 percent in the year after, and 5 percent in the third year after the claim. The regular premium is, in general, about twice the regular deductible amount, so an

[45] This assumption is also supported by the following observation. In an informal survey we conducted among our colleagues, all of them were aware of a deductible in their auto insurance policy, but less than 20 percent knew its level. This does not imply that 80 percent of our colleagues did not pay attention to their deductible choice at the time the choice was made. It does imply, however, that their behavior cannot depend on the deductible amount.

upper bound for the dynamic costs is about 70%. In most cases the actual dynamic costs are much lower than this upper bound, as the dynamic costs of, say, the first claim within a year are minimal. In addition, an individual can always opt out of the contract and switch to a different insurance provider. This is likely to reduce her dynamic costs because in Israel, unlike in the US and in many other countries, there is no public record for past claims. Therefore, insurance providers can take full advantage of past records only for their past customers. For this reason, new customers will, of course, face higher premia than existing ones, but the premium increase would not be as high as it would have been with the old insurance provider. This is due to the presence of "innocent" new customers, who are pooled together with the switchers (see also Cohen, 2003).[46] Using this 70% as a conservative upper bound, we can repeat a similar exercise to find out that about 93% of those claims filed by individuals with a low deductible were higher than 1.7 times the regular deductible level. While this is not negligible, it applies for only a tiny fraction of the individuals. For the vast majority of them, the 99% figure is the relevant one. Therefore, ex-post moral hazard is unlikely to play a major role in this setting, and one can abstract from the loss distribution and focus on claim rates, as we do in this paper.[47] Finally, we should note that, as with driving behavior, to the extent that this assumption slightly biases our results, it should do so by making the choice of a low deductible slightly less attractive than we estimate it to be, thus implying that individuals may be slightly more risk averse than our estimates suggest.

Incomplete information by individuals We assume through most of the paper that individuals have perfect information about their risk types λ_i. Note, first, that this is a stronger assumption than we need. Under expected utility framework, utility is linear in probabilities, so all we need is that individuals' *expected* risk rate is the true one, i.e. $\widehat{\lambda}_i \equiv E(\widetilde{\lambda}_i|I_i) = \lambda_i$ where $\widetilde{\lambda}_i$ is individual i's perceived risk rate, and I_i is individual i's information at the time of the coverage choice. Namely, individuals may be uncertain about their risk type, but their point estimate is correct. Still, it is reasonable to argue that this is a rather extreme assumption. There are several channels through which incomplete information may operate. Let us consider two such cases. First, suppose that individuals are correct, but only on average, i.e. that $\widehat{\lambda}_i = \lambda_i + \epsilon_i$ where $E(\epsilon_i) = 0$. The intuition for this case is similar to an "errors in variables" model, and in principle will result in an even less important role for adverse selection. Given that we find relatively little role for adverse selection, this bias will not change this conclusion. This may be even more pronounced if $Corr(\lambda_i, \epsilon_i) < 0$, which reflects a reasonable assumption of "reversion to the mean," i.e. that individuals' estimates of their risk types are some weighted average of their true risk types and the average risk type of individuals who are similar (on observables) to them. The conclusion may go in the other way

[46] New customers may voluntarily report their claim history to their new insurance provider. Voluntary disclosure of past claims is, as may be expected, not truthful. Our data suggest an unconditional claim rate of 24.53% in our sample population. Our data on claim history, as voluntarily disclosed by the same indivduals, suggest a claim rate of 6.04%, which is four times lower.

[47] This last statement may not be as clean once we take into account the choice of "high" and "very high" deductibles, which are at much higher levels. This is one additional reason to focus only on the choice between "low" and "regular" deductibles.

only if the mistakes go in the other direction, according to which individuals who are riskier than average believe that they are even more risky than they truly are. This, we believe, is less likely.

Finally, as discussed in the previous section, Table 8 reports the results for experienced drivers only, who are likely to have better information about their risk rates. The fact that the results do not change much may suggest that the main results are not highly sensitive to heterogeneity in information. In addition, Table 9 reports the estimation of the model under a particular incomplete information assumption, and also suggests that the main qualitative results are robust to the information structure.

Additional cost of an accident Our model assumes that, in the event of an accident, the only incurred costs are those associated with the deductible payment. In practice, however, other transaction costs may be associated with an accident, such as the time spent for appraisal of the damage, the costs associated with renting a replacement car for the duration of a repair, etc. Such costs could be readily incorporated into the model; one can think about them as an additional (mandatory) deductible. To illustrate, we assume that these costs are known in advance and are given by a constant c (which could, in principle, vary with each individual). Since c will not vary with the chosen level of deductible, it will not affect the value Δd and will only enter the empirical model through its effect on $\overline{d}$. In particular, equation (7) will change to

$$r \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\overline{d} + c} \tag{23}$$

and everything else will remain the same.

This implies that, in principle, such costs will have no effect on the results of the counterfactual exercise, which are still valid. The costs will, however, affect the interpretation of the estimates of risk aversion. In particular, instead of the distribution of r we will now be estimating the distribution of $r\frac{\overline{d}+c}{\overline{d}}$, so the reported estimates of the coefficient of absolute risk aversion will be biased upwards. The magnitude of the bias depends on the size of these transaction costs c compared to the average deductible $\overline{d}$. If c is relatively small, the bias is negligible. If, however, c is as big as the (average) deductible level, all our reported estimates of the level of risk aversion should be divided by two (but the coefficients on observables, which are semi-elasticities, will not change). The intuition would be similar, but more involved, if c varies across individuals but not proportionally to $\overline{d}$. In the absence of data about transaction costs, all one can do is introspect and use priors for the relative importance of c.

Sample selection Given that the company is a somewhat non-standard insurance provider and that it is new in the market, it is potentially more likely to attract individuals who are more likely to experiment with new ways to do business, and may be in general less risk averse than the general population. In Table 10 we compare the demographics of our sample of policyholders with those of the general Israeli population. This comparison reflects a similar intuition: compared with the general population our average policyholder is somewhat younger, more educated, more likely to be a male, and less likely to be married or an immigrant. This direction of selection may also apply to unobserved risk preferences, thereby making our policyholders, on average, less risk averse than a representative individual. This may suggest that the level of risk aversion that we find can be viewed as a lower bound on the level of risk aversion in the general population.

One potential way to model sample selection is to allow for an additional outside option to be selected. For the vast majority of the individuals we observe, the outside option is to purchase similar insurance from competing insurance agencies. Unfortunately, data on the structure of competing contracts, their prices, and the way they vary by individual characteristics are unavailable. This makes us uncomfortable to try to model sample selection, as results from any such model will be driven by our assumptions rather than by meaningful variation in the data. Therefore, we choose to report the results for the sampled, potentially selected population. The results are still meaningful for two reasons. First, as mentioned before, this is a large population, accounting for about seven percent of all drivers in Israel. Second, to the extent that our estimates suggest higher levels of risk aversion than previously estimated and that the sample selection is likely to bias these estimates downwards, the results are still highly informative.

One should note that the potential sample selection discussed above has no direct impact for the counterfactual exercise, which should be taken with respect to the population at hand. There is, however, a related selection problem, which may affect the counterfactuals. In our counterfactual exercise we assume that as the company, say, increases the low deductible level, consumers switch to higher coverage, but remain with the same company. If increasing the low deductible makes individuals switch to a different company, the benefits from increasing the low deductible would be lower. We believe that this is not a major problem because of the significant cost advantage the company enjoyed and the evidence in the literature (discussed in the beginning of Section 2) emphasizing that the choice of direct insurers is driven, to a large extent, by non-monetary "amenities." This makes it reasonable to think about the choice we analyze as a nested decision problem: which type of company to choose, and then which deductible level. This makes our counterfactual analysis valid. Finally, one should note that the two potential selection problems cannot both be important at the same time. The first may be important only if direct insurers are perceived to be very different from traditional insurers, while the second is important only if all insurers are the same, so competition is primarily channeled through the financial parameters of the contracts.

Deviations from expected utility theory Throughout the paper we restrict attention to expected utility maximizers. Despite much evidence in the literature against some of the predictions of expected utility theory, it still seems to us the most natural benchmark to specify, and one which facilitates comparison to previous studies. It is important to note that expected utility theory is assumed; it is not and cannot be tested within our framework. Given our cross-sectional analysis, which, in principle, allows flexible forms of unobserved heterogeneity in risk preferences, there are no testable restrictions imposed by expected utility theory. We should also note that much (but not all) of the documented evidence against expected utility theory arises with extreme risk probabilities, which are close to zero or one. Our data (and our estimates) are based on risk probabilities which are roughly in the range of $0.1 - 0.35$ (see Figure 4). Over this range, expected utility seems to perform better. Finally, it is important to stress two points. First, at the conceptual level, it is straightforward to use an alternative theory of decisions under uncertainty. If, conditional on objective risk, individuals vary in a single dimension, the same conceptual model and empirical strategy can be applied. All one needs to do is to specify the parameter over which decisions vary, and construct an indifference set in the space of the specified parameter and (objective) risk types, similar to the one presented in Figure 3. Second, any alternative model of decisions under uncertainty would require us to take an even stronger view regarding the parameterized objective function. For example, prospect theory (Kahneman and Tversky, 1979) would require us to parameterize not only the curvature of individuals' utility functions, but, in addition, their reference points, for which there is no natural choice in our context. Similar issues will arise if we tried to apply decision weights (Tversky and Wakker, 1995) or measures of, say, over-confidence with respect to driving ability.

6 Summary and Concluding Remarks

The paper makes two separate contributions. First, from a methodological standpoint, we lay out a conceptual framework through which one can formulate and estimate a demand system for individually-customized contracts. The key data requirements for this approach are contract choices, individual choice sets, and ex-post risk realizations. Since such data may be available in other settings, the methodological framework may be useful to uncover structural parameters in such settings. As an example, one could consider annuity data and use guarantee-period choices and mortality data to identify between heterogeneity in risk (mortality) and in preferences, such as bequest motives. Similarly, one could consider loan data and use loan-duration choices and default data to identify between heterogeneity in risk (default) and in credit constraints.

Second, from an economic standpoint, we provide a new set of estimates for the degree and heterogeneity of (absolute) risk aversion. Our results can be summarized by the following four findings. First, the average coefficient of absolute risk aversion we estimate is higher than other estimates available in the literature. Second, we find important relationships between individual characteristics and risk aversion. In particular, we find positive correlation between the degree of risk avesion and various measures of income or wealth. Third, we find that unobserved risk aversion is positively correlated with unobserved risk. Finally, we find that unobserved heterogeneity in risk aversion is greater than unobserved heterogeneity in risk. The last two findings have important implications for the design and pricing of optimal insurance contracts.

Many econometric applications structurally estimate utility parameters. Often, however, these estimated utility parameters are context dependent (e.g. the willingness to pay for an extra horse power of a car), and therefore are not interesting *per se*. Rather, we are interested in the implication of their estimated values for pricing, substitution patterns, and welfare. Our estimates of risk aversion also serve a similar role; they inform us about optimal pricing and welfare in the auto insurance market. In addition, however, we might be interested in their values directly, as they may help to explain choices made by individuals in other risky contexts.

It is natural to ask to what extent these parameters are relevant in other contexts. This is essentially an empirical question, which can be answered only by estimating risk aversion parameters for a variety of bet sizes and in a variety of contexts. Since isolating risk preferences in many contexts is hard, such exercises are rare (providing the main motivation for the current work), leaving us with no definite answer for the scope of markets for which our estimates may be relevant. On one hand, Rabin (2000) and Rabin and Thaler (2001) argue that different decisions in life are taken in different contexts, and therefore may be subject to different parameters in the utility function (or, equivalently, to different "relevant wealth levels"). According to this view, one should not extrapolate our estimates to other contexts. On the other hand, classical theory suggests that each individual has one value function over her life-time wealth, so all risky decisions take into account the same value function and are therefore subject to the same risk preferences. Our view is somewhere in between. We are more comfortable with extrapolation of our risk aversion estimates to setups which are "closer" to the auto insurance market context in which these parameters are estimated. When doing so, one should assess various factors over which contexts may differ. Such factors include the size of the bets and different factors that may change informational and behavioral effects, such as default options, framing, and rarity of the events considered. Finally, selection of participants into a particular market may be an important factor. Given our estimate of large heterogeneity of unobserved risk preferences, one may expect strong selection on unobservables in a wide range of voluntary markets, making our estimates more representative, perhaps, for markets in which participation is either mandatory or driven by other factors, which are unlikely to be correlated with risk preferences.

Selection into the market may be the primary reason why our estimates of risk aversion are generally higher than comparable estimates found in the literature (in, say, Gertner, 1993; Metrick, 1995; and Jullien and Salanie, 2000). Some economists consider our average estimates of risk aversion as surprisingly high. By introspection, we think that an average individual who is indifferent about a fifty-fifty lottery of gain $100 lose $61 is not unreasonable. In fact, many economists believe that high static risk aversion is a good candidate for explaining the high equity premium.[48] To the extent that one is comfortable extrapolating our risk aversion estimates to other settings, they provide some support to this view.

Our second set of findings concerns the way risk aversion relates to observable characteristics. These results may be more robust across different settings. Even if the magnitude of risk aversion may be sensitive to context, the way it varies with gender, wealth, or age may be less context dependent. Thus, for example, our finding that women are more risk averse than men may generalize across a variety of situations. Other findings suggest that the estimated coefficient of risk aversion increases with observables that are related to income and wealth. As already discussed, this does not necessarily imply that we reject the widely held belief of the decreasing absolute risk aversion

[48] See Mehra and Prescott (1985), and Kocherlacota (1996) for a more recent review.

property. The finding is more plausibly rationalized by an underlying positive correlation between wealth and risk aversion across individuals. Testing between these two alternatives would require different data, with a panel structure and exogenous shocks to wealth. In a cross-section, the underlying relationship between wealth and risk preferences can go either way. While lower risk aversion may be associated with higher propensity to become an entrepreneur, and thereby have higher wealth, it may also be associated with lower propensity to save or invest in education, affecting wealth the other way. Therefore, one important message of this finding is that accounting for heterogeneity in preferences may be important, as representative consumer models may provide misleading interpretations for otherwise natural results.

The last two sets of findings concern the relative heterogeneity in risk aversion and that of risk, and the correlation between risk and risk aversion. In contrast to the preceding discussion, these findings may be sensitive to the market context. While measures of risk aversion may extend across contexts, the measure of risk is specific to the application we study. Thus, any relationship between risk and risk aversion may change once risk takes other forms. First, within the auto insurance market, one could measure accident risk in various ways. We use claim propensity, as this is the relevant measure for the decision at hand. Other measures of risk, such a risk per mile travelled, injury propensity, or expected value of the claim may relate differently to risk aversion, as they depend on a different set of latent variables. Moreover, risk in the auto insurance market may be conceptually different from risk in other markets. In many insurance markets, such as life insurance, health insurance, and long term care insurance, risk is independent across individuals, and is therefore primarily driven by individual characteristics and behavior. In auto insurance, much of the risk depends on coordination among drivers, and therefore may be more related to relative (rather than absolute) individual characteristics and behavior.

Thus, although our finding of a strong positive correlation between risk and risk aversion may be somewhat surprising, we do not view it as an evidence against other findings of negative correlation between risk and risk aversion in other contexts, such as long term care insurance (Finkelstein and McGarry, 2003). The positive correlation we find is also consistent with the fact that the reduced-form bivariate Probit tests for adverse selection in our data provide evidence for adverse selection (Cohen, 2005), while similar reduced-form tests in other contexts do not. As noted by Finkelstein and McGarry (2003), positive correlation does make these tests more likely to conclude that adverse selection exists.

Finally, we find that unobserved heterogeneity in risk preferences is more important than heterogeneity in risk. This general finding seems consistent with the general message of the recent influential literature on adverse selection (Chiappori and Salanie, 2000; Finkelstein and McGarry, 2003; Chiappori et al., forthcoming). It might be driven by the casual evidence that insurance companies exert much effort and resources in collecting consumer data, which are informative about risk classification, but not much data about risk preferences.[49] We illustrate the empirical importance

[49] This choice of data collection efforts may be justified if it is easier for such firms to price discriminate based on risk, but it is harder to price discriminate based on preferences. Without cost-based (i.e. risk-based) justification for prices, price discrimination may lead, for example, to consumer backlash.

of our findings for the analysis of optimal contracts in auto insurance. The presence of more than one dimension of unobserved heterogeneity may dramatically change the nature of these contracts, and the nature of the observed relationship in the data. Theory is still not fully developed for such multi-dimensional screening problems, as it typically requires a small number of types (Landsberger and Meilijson, 1999), restricts the two dimensions to be independent of each other (Rochet and Stole, 2002), or assumes that the number of instruments available to the monopolist is not smaller than the dimension of unobserved heterogeneity (Matthews and Moore, 1987). Armstrong (1999), who analyzes optimal regulatory contracts in the presence of both cost and demand uncertainties, may be the closest theoretical work to the framework suggested here. It cannot be directly applied, however, as it uses simplifying linearity assumptions, which would be hard to impose in the current context. Our results indicate that many applications can benefit from extending the theory to include the more general case, such as the one analyzed here. Such a theory may also serve as a guide for using supply-side moment conditions in this context. Our counterfactual analysis in Section 4.3 is a very preliminary start in this direction.

References

[1] Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutstrom (2004), "Preference Heterogeneity in Experiments: Comparing the Field and Lab," mimeo, University of Central Florida.

[2] Armstrong, Mark (1999), "Optimal Regulation with Unknown Demand and Cost Functions," *Journal of Economic Theory*, 84, 196-215.

[3] Arnott, Richard, and Joseph E. Stiglitz (1988), "Randomization with Asymmetric Information," *Rand Journal of Economics*, 29(3), 344-362.

[4] Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro (1997), "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 112(2), 537-579.

[5] Beetsma, Roel M. W. J., and Peter C. Schotman (2001), "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo," *The Economic Journal*, 111, 821-848.

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

[7] Cicchetti, Charles J., and Jeffrey A. Dubin (1994), "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy*, 102(1), 169-186.

[8] Chetty, Raj (2004), "Labor Supply and Risk Aversion: A Calibration Theorem," mimeo, UC Berkeley.

[9] Chiappori, Pierre-Andre, and Bernard Salanie (2000), "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108(1), 56-78.

[10] Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie (forthcoming), "Asymmetric Information in Insurance: General Testable Implications," *Rand Journal of Economics*.

[11] Chiappori, Pierre-Andre, and James Heckman (2000), "Testing for Moral Hazard on Dynamic Insurance Data: Theory and Econometric Tests," mimeo, University of Chicago.

[12] Cohen, Alma (2003), "Profits and Market Power in Repeat Contracting: Evidence from the Insurance Market," mimeo, NBER.

[13] Cohen, Alma (2005), "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market," *Review of Economics and Statistics*, 87(2), forthcoming.

[14] Cohen, Alma, and Liran Einav (2003), "The Effect of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," *Review of Economics and Statistics*, 85(4), 828-843.

[15] Cooper, Russell, and Beth Hayes (1987), "Multi-period Insurance Contracts," *International Journal of Industrial Organization*, 5(2), 211-231.

[16] Cummins, J. David, and Jack L. Van Derhei (1979), "A Note on the Relative Efficiency of Property-Liability Insurance Distribution System," *Bell Journal of Economics*, 10, 709-719.

[17] Damien, Paul, John Wakefield, and Stephen Walker (1999), "Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables," *Journal of the Royal Statistical Society B*, 61(2), 331-344.

[18] D'arcy, Stephen P., and Neil A. Doherty (1990), "Adverse Selection, Private Information, and Lowballing in Insurance Markets," *Journal of Business*, 63(2), 145-164.

[19] De Meza, David, and David C. Webb (2001), "Advantegeous Selection in Insurance Markets," *Rand Journal of Economics*, 32(2), 249-262.

[20] Devroye, Luc (1986), *Non-Uniform Random Variate Generation*, Springer-Verlad, New York.

[21] Dionne, Georges, and Neil A. Doherty (1994), "Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from Insurance Markets," *Journal of Political Economy*, 102(2), 209-235.

[22] Dionne, Georges, and Pierre Lasserre (1985), "Adverse Selection, Repeated Insurance Contracts, and Announcement Strategy," *Review of Economics Studies*, 52(4), 719-723.

[23] Dionne, Georges, and Charles Vanasse (1992), "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information," *Journal of Applied Econometrics*, 7(2), 149-165.

[24] Donkers, Bas, Bertrand Melenberg, and Arthur van Soest (2001), "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach," *The Journal of Risk and Uncertainty*, 22(2), 165-195.

[25] Dreze, Jacques H. (1981), "Inferring Risk Tolerance from Deductibles in Insurance Contracts," *The Geneva Papers on Risk and Insurance*, 20, 48-52.

[26] Evans, William N., and W. Kip Viscusi (1991), "Estimation of State-Dependent Utility Functions Using Survey Data," *Review of Economics and Statistics*, 73(1), 94-104.

[27] Finkelstein, Amy, and Kathleen McGarry (2003), "Private Information and Its Effect on Market Equilibrium: New Evidence from Long-Term Care Insurance," mimeo, Harvard University.

[28] Finkelstein, Amy, and James Poterba (2004), "Adverse Selection in Insurance Markets: Policyholder Evidence from the UK Annuity Market," *Journal of Political Economy*, forthcoming.

[29] Friend, Irwin, and Marshall E. Blume (1975), "The Demand for Risky Assets," *American Economic Review*, 65(5), 900-922.

[30] Gertner, Robert (1993), "Game Shows and Economic Behavior: Risk-Taking on "Card Sharks"," *Quarterly Journal of Economics*, 108(2), 507-521.

[31] Geweke, John (1991), "Efficient Simulation from the Multivariate Normal and Student-t Distributions subject to Linear Constraints," in E. M. Keramidas (ed.) Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, 571-578.

[32] Hartog, Joop, Ada Ferrer-i-Carbonell, and Nicole Jonker (2002), "Linking Measured Risk Aversion to Individual Characteristics," *Kyklos*, 55, 3-26.

[33] Heckman, James J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153-162.

[34] Hendel, Igal, and Alessandro Lizzeri (2003), "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *Quarterly Journal of Economics*, 118(1), 299-327.

[35] Holt, Charles A., and Susan K. Laury (2002), "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644-1655.

[36] Israel, Mark (2005), "Where is All the Hidden Information Hiding? Evidence from Automobile Insurance Panel Data," mimeo, Northwestern University.

[37] Jullien, Bruno, and Bernard Salanie (2000), "Estimating Preferences under Risk: The Case of Racetrack Bettors," *Journal of Political Economy*, 108(3), 503-530.

[38] Jullien, Bruno, Bernard Salanie, and Francois Salanie (2003), "Screening Risk-averse Agents under Moral Hazard: Single-Crossing and the CARA Case," mimeo, Universite de Toulouse.

[39] Kachelmeier, Steven J., and Mohamed Shehata (1992), "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82(5), 1120-1141.

[40] Kahneman, Daniel, and Amos Tversky (1979), "Prospect Theory: An Analysis of Decisions Under Risk," *Econometrica*, 47(2), 263-292.

[41] Kocherlakota, Narayana R. (1996), "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42-71.

[42] Kunreuther, Howard, and Mark V. Pauly (1985), "Market Equilibrium with Private Knowledge: An Insurance Example," *Journal of Public Economics*, 26(3), 269-288.

[43] Landsberger, Michael, and Isaac Meilijson (1999), "A General Model of Insurance under Adverse Selection," *Economic Theory*, 14, 331-352.

[44] Matthews, Steven, and John Moore (1987), "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening," *Econometrica*, 55(2), 441-467.

[45] Mehra, Ranjish, and Edward C. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145-161.

[46] Metrick, Andrew (1995), "A Natural Experiment in "Jeopardy!"," *American Economic Review*, 85(1), 240-253.

[47] Palacios-Huerta, Ignacio, Roberto Serrano, and Oscar Volij (2003), "Rejecting Small Gambles Under Expected Utility," mimeo, Brown University.

[48] Puelz, Robert, and Arthur Snow (1994), "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market," *Journal of Political Economy*, 102, 236-257.

[49] Rabin, Matthew (2000), "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68(5), 1281-1292.

[50] Rabin, Matthew, and Richard H. Thaler (2001), "Anomalies: Risk Aversion," *Journal of Economic Perspectives*, 15(1), 219-232.

[51] Rochet, Jean-Charles, and Lars A. Stole (2002), "Nonlinear Pricing with Random Participation," *Review of Economic Studies*, 69(1), 277-311.

[52] Rothschild, Michael, and Joseph E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90(4), 629-649.

[53] Rubinstein, Ariel (2001), "Comments on the Risk and Time Preferences in Economics," mimeo, Tel Aviv University.

[54] Saha, Atanu (1997), "Risk Preference Estimation in the Nonlinear Mean Standard Deviation Approach," *Economic Inquiry*, 35(4), 770-782.

[55] Smart, Michael (2000), "Competitive Insurance Markets with Two Unobservables," *International Economic Review*, 41(1), 153-169.

[56] Smith, Vernon L., and James M. Walker (1993), "Rewards, Experience, and Decision Costs in First Price Auctions," *Economic Inquiry*, 31(2), 237-244.

[57] Tanner, M. A.,and W. H. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of The American Statistical Association,* 82, 528-549.

[58] Tversky, Amos, and Peter Wakker (1995), "Risk Attitudes and Decision Weights," *Econometrica*, 63(6), 1255-1280.

[59] Villeneuve, Bertrand (2003), "Concurrence et Antisélection Multidimensionnelle en Assurance," *Annales d'Economie et Statistique*, 69, 119-142.

[60] Viscusi, W. Kip, and William N. Evans (1990), "Utility Functions that Depend on Health Status: Estimates and Economic Implications," *American Economic Review*, 80(3), 353-374.

[61] Watt, Richard (2002), "Defending Expected Utility Theory," *Journal of Economic Perspectives*, 16(2), 227-230.

Appendix

A Gibbs Sampler

In this appendix we describe the setup of the Gibbs Sampler that we use to estimate the model. One of the main advantages of the Gibbs Sampler is its ability to allow for data augmentation of latent variables (Tanner and Wong, 1987). In our context, this amounts to augmenting the individual-specific risk aversion and risk type, namely $\{\lambda_i, r_i\}_{i=1}^n$ as additional parameters.

We can write the model as follows:

$$\ln \lambda_i = x_i'\beta + \varepsilon_i \tag{24}$$

$$\ln r_i = x_i'\gamma + \upsilon_i \tag{25}$$

$$choice_i = \begin{cases} 1 & if \quad r_i > r_i(\lambda) \\ 0 & if \quad r_i < r_i(\lambda) \end{cases} \tag{26}$$

$$claims_i \sim Poisson(\lambda_i t_i) \tag{27}$$

$$\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{pmatrix}\right] \tag{28}$$

Let $\delta \equiv \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$, $\Sigma \equiv \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{pmatrix}$, $X \equiv \begin{pmatrix} \begin{pmatrix} x_1 \\ : \\ x_n \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} x_1 \\ : \\ x_n \end{pmatrix} \end{pmatrix}$, $y \equiv \begin{pmatrix} \lambda \\ r \end{pmatrix}$, and $u_i \equiv \begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix}$. The set of parameters for which we want to have a posterior distribution is given by $\theta = \{\delta, \Sigma, \{u_i\}_{i=1}^n\}$. The prior specifies that $\{\delta, \Sigma\}$ are independent of $\{u_i\}_{i=1}^n$. $\{\delta, \Sigma\}$ have a conventional diffuse prior. We adopt a hierarchical prior for $\{u_i\}_{i=1}^N$:

$$\{u_i\}_{i=1}^n | \Sigma \overset{iid}{\sim} N(0, \Sigma) \tag{29}$$

$$\Sigma^{-1} \sim Wishart_2(a, Q) \tag{30}$$

so conditional on all other parameters (and on the data, which have no effect in this case), we have:

$$\Sigma^{-1} | \delta, \{u_i\}_{i=1}^n \sim Wishart_2\left(a + n - k, \left(Q^{-1} + \sum_i u_i u_i'\right)^{-1}\right) \tag{31}$$

and

$$\delta|\Sigma, \{u_i\}_{i=1}^n \sim N\left((X'X)^{-1}(X'y), \Sigma^{-1} \otimes (X'X)^{-1}\right) \tag{32}$$

For Σ^{-1} we use a convenient diffuse prior, i.e. $a = 0$ and $Q^{-1} = 0$.

The part of the Gibbs Sampler which is less standard in this case involves the sampling from the conditional distribution of the augmented parameters, $\{u_i\}_{i=1}^n$. Each individual is independent of the others, so conditional on the other parameters, it does not depend on other individuals' augmented data. Thus, all we need to describe is the conditional probability of u_i. Note that conditional on δ we have $\varepsilon_i = \ln\lambda_i - x_i'\beta$ and $\upsilon_i = \ln r_i - x_i'\gamma$ so we can instead focus on sampling from the posterior distribution of λ_i and r_i. These posterior distributions are as follows:

$$\Pr\left(r_i|\gamma,\beta,\Sigma,\lambda_i,data\right) \propto \begin{cases} \phi\left[\ln r_i, x_i'\gamma + \rho\frac{\sigma_r}{\sigma_\lambda}(\lambda_i - x_i'\beta), \sqrt{\sigma_r^2(1-\rho^2)}\right] & if\ choice_i = I(r_i < r_i(\lambda)) \\ 0 & if\ choice_i \neq I(r_i < r_i(\lambda)) \end{cases} \tag{33}$$

and

$$\Pr\left(\lambda_i|\gamma,\beta,\Sigma,r_i,data\right) \propto$$

$$\propto \begin{cases} p(\lambda_i, claims_i, t_i)\phi\left[\ln\lambda_i, x_i'\beta + \rho\frac{\sigma_\lambda}{\sigma_r}(r_i - x_i'\gamma), \sqrt{\sigma_\lambda^2(1-\rho^2)}\right] & if\ choice_i = I(r_i < r_i(\lambda)) \\ 0 & if\ choice_i \neq I(r_i < r_i(\lambda)) \end{cases} \tag{34}$$

where $p(x, claims, t) = x^{claims}\exp(-xt)$ is proportional to the probability density function of the Poisson distribution, and $\phi(x,\mu,\sigma) = \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ is proportional to the Normal probability density function.

The posterior for $\ln r_i$ is a truncated normal, for which we use a simple "invert cdf" sampling (Devroye, 1986).[50] The posterior for $\ln\lambda_i$ is more tricky. We use a "slice sampler" to do so (Damien et al., 1999). The basic idea is to rewrite $\Pr(\lambda_i) = b_0(\lambda_i)b_1(\lambda_i)b_2(\lambda_i)$ where $b_0(\lambda_i)$ is truncated normal distribution, and $b_1(\lambda_i)$ and $b_2(\lambda_i)$ are defined below. We can then augment the data with two additional variables, u_i^1 and u_i^2, which (conditional on λ_i) are distributed uniformly on $[0, b_1(\lambda_i)]$ and $[0, b_2(\lambda_i)]$, respectively. Then we can write $\Pr\left(\lambda_i, u_i^1, u_i^2\right) = b_0(\lambda_i)b_1(\lambda_i)b_2(\lambda_i)\frac{I(0\le u_i^1 \le b_1(\lambda_i))}{b_1(\lambda_i)}\frac{I(0\le u_i^2\le b_2(\lambda_i))}{b_2(\lambda_i)} = b_0(\lambda_i)I(0 \le u_i^1 \le b_1(\lambda_i))I(0 \le u_i^2 \le b_2(\lambda_i))$. Using this form we have that $b_1(\ln\lambda_i) = \lambda_i^{claims_i} = (\exp(\ln\lambda_i))^{claims_i}$ and $b_2(\ln\lambda_i) = \exp(-\lambda_i t_i) = \exp(-t_i\exp(\ln\lambda_i))$. Because $b_1(\cdot)$ and $b_2(\cdot)$ are both monotone functions, conditional on u_{1i} and u_{2i} this just means that $b_1^{-1}(u_{1i}) = \frac{\ln u_{1i}}{claims_i}$ is a lower bound for $\ln\lambda_i$ (for $claims_i > 0$) and that $b_2^{-1}(u_{2i}) = \ln(-\ln u_{2i}) - \ln t_i$ is an upper bound for $\ln\lambda_i$. Thus, we can just sample λ_i from a truncated normal distribution, after we modify the bounds according to u_i^1 and u_i^2.

In the end of Section 4 we add to the model incomplete information of individuals about their own types. Individuals' type are fixed over their lifetime, and individuals are Bayesian and update

[50] Let $F(x)$ be the cumulative distribution function. The "invert cdf" sampling draws from this distrubtion by drawing u from $[0,1]$ and computing $F^{-1}(u)$. In principle, one can use the sampling procedure suggested by Geweke (1991), which avoids computing F^{-1} and therefore is more efficient. It turns out, however, that vectorizing the algorithm is much easier when we use Devroye (1986). The vectorization entails enormous computational benefits when coded in Matlab.

their own type by their claim experience. Since expected utility is linear in claim probabilities, only individuals' ex-post mean will affect their coverage choices. If individuals have a prior which follows a $Gamma(\alpha, \beta)$ distribution,[51] their posterior mean is given by $\frac{c_i+\alpha}{l_i+\frac{1}{\beta}}$ where c is the number of claims historically filed and l_i is the individual's driving experience (license years). Let $\widehat{\lambda}_i$ denote the posterior mean. The assumptions imply that $\widehat{\lambda}_i(l_i + \frac{1}{\beta}) - \alpha$ is distributed $Poisson(\lambda_i l_i)$. The rest of the model is as before, with $\widehat{\lambda}_i$ used instead of λ_i to explain the coverage choice. Thus, to implement it within the Gibbs sampler, we augment $\widehat{\lambda}_i$ as well. The conditional distribution for $r_i|\lambda_i, \widehat{\lambda}_i$ is as before, with $\widehat{\lambda}_i$ (rather than λ_i) affecting the truncation point. The conditional distribution for $\widehat{\lambda}_i|\lambda_i, r_i$ is a linear transformation of a truncated Poisson, with the truncation point is the same as the one used above for λ_i. Finally, the conditional distribution for $\lambda_i|\widehat{\lambda}_i, r_i$ is, as before, of an unknown form. Fortunately, however, the assumptions make it very similar to the one before, with the following modifications. First, it is not truncated. Second, $\widehat{\lambda}_i$ provides information on λ_i. In particular, it can be written as:

$$\Pr\left(\lambda_i|\gamma, \beta, \Sigma, \widehat{\lambda}_i, r_i, data\right) \propto$$

$$\propto p(\lambda_i, \widehat{\lambda}_i(l_i + \frac{1}{\beta}) - \alpha, l_i)p(\lambda_i, claims_i, t_i)\phi\left[\ln\lambda_i, x_i'\beta + \rho\frac{\sigma_\lambda}{\sigma_r}(r_i - x_i'\gamma), \sqrt{\sigma_\lambda^2(1-\rho^2)}\right] \tag{35}$$

Because the two first elements follow a Poisson process, however, it is proportional to $p(x, claims, t) = x^{claims_i + c_i}\exp(-x(t_i + l_i))$, making it very similar to the form of the benchmark model.

B Variable Definitions

Below we describe the variables which may not be self-explanatory:

- Education - "Technical" education refers to post high school education, which does not result in an academic degree.
- Emigrant - A dummy variable which is equal to 1 if the individual was not born in Israel.
- Car value - Current estimated "blue book" value of the car.
- License years - Number of years since the individual obtained driving license.
- Good driver - A dummy variable which is equal to 1 if the individual is classified as a good driver. The classification is made by the company, based on the other observables, and suggests that the individual is likely to be a low-risk driver. We do not know the exact functional form for this classification. One can view this as an informative non-linear functional form of the other observables already in the regressions.

[51] Note that the Gamma assumption is similar, but not identical, to the Loglinear distribution we use for estimation. As will be clear below, the benefit of this slight internal inconsistency is very attractive computationally, for the construction of the Gibbs sampler.

- "Any Driver" - A dummy variable which is equal to 1 if the policy stipulates that any driver can drive the car. If it does not stipulate it, the car is insured only if the policy holder (and sometimes his/her spouse) drives the car.
- Secondary car - A dummy variable which is equal to 1 if the car is not the main car in the household.
- Business use - A dummy variable which is equal to 1 if the policy holder uses the car for business.
- Commercial car - A dummy variable which equal to 1 if the car is defined as a commercial vehicle (e.g. pick-up truck).
- Estimated mileage - Predicted annual mileage (in kilometers) by the policy holder. The company does not use this variable for pricing, as it is believed to be unreliable.
- History - The number of years (up to 3) prior to the starting date of the policy for which the policy holder reports his/her past claim history.
- Claim history - The number of claims per year for the policy holder over the 3 (sometimes less) years prior to the starting date of the policy.
- Young driver - For drivers below the age of 25, the policy holder has to separately report the details of the young driver (which may be the policy holder or someone else).
- Company year - Year dummies, which span our five-year observation period. The first year dummy is equal to 1 for policies started between 11/1/1994 and 10/31/1995, the second year dummy is equal to 1 for policies started between 11/1/1995 and 10/31/1996, and so forth.

Table 1: Descriptive Statistics - Covariates

	Variable		Obs	Mean	Std. Dev.	Min	Max
Demographics:	Age		105,800	41.14	12.37	18.06	89.43
	Female		105,800	0.32	0.47	0	1
	Family	Single	105,800	0.143	0.35	0	1
		Married	105,800	0.779	0.42	0	1
		Divorced	105,800	0.057	0.23	0	1
		Widower	105,800	0.020	0.14	0	1
		Refused to Say	105,800	0.001	0.04	0	1
	Education	Elementary	105,800	0.016	0.12	0	1
		High School	105,800	0.230	0.42	0	1
		Technical	105,800	0.053	0.22	0	1
		Academic	105,800	0.233	0.42	0	1
		No Response	105,800	0.468	0.50	0	1
	Emigrant		105,800	0.335	0.47	0	1
Car Attributes:	Value (current NIS)[a]		105,800	66,958.41	37,376.76	4,000	617,000
	Car Age		105,800	3.91	2.94	0	14
	Commercial Car		105,800	0.083	0.28	0	1
	Engine (cc)		105,800	1,567.94	384.68	700	5,000
Driving:	License Years		105,798	18.18	10.08	0	63
	Good Driver		105,800	0.548	0.50	0	1
	"Any Driver"		105,800	0.257	0.44	0	1
	Secondary Car		105,800	0.151	0.36	0	1
	Business Use		105,800	0.082	0.27	0	1
	Estimated Mileage (km)		60,422	14,031.09	5,890.50	1,000	32,200
	History		105,800	2.847	0.61	0	3
	Claim History		105,800	0.060	0.15	0	2.00
Young Driver:	Age	17-19	105,800	0.029	0.17	0	1
		19-21	105,800	0.051	0.22	0	1
		21-24	105,800	0.089	0.29	0	1
		>24	105,800	0.022	0.15	0	1
	Experience	<1	105,800	0.042	0.20	0	1
		1-3	105,800	0.071	0.26	0	1
		>3	105,800	0.079	0.27	0	1
	Gender	Male	105,800	0.113	0.32	0	1
		Female	105,800	0.079	0.27	0	1
Company Year:	First year		105,800	0.207	0.41	0	1
	Second year		105,800	0.225	0.42	0	1
	Third year		105,800	0.194	0.40	0	1
	Fourth year		105,800	0.178	0.38	0	1
	Fifth year		105,800	0.195	0.40	0	1

[a] The average exchange rate throughout the sample period was approximately 1 US dollar per 3.5 NIS, starting at 1:3 in late 1994 and reaching 1:4 in late 1999.

Table 2A: Descriptive Statistics - Menus, Choices, and Outcomes

	Variable		Obs	Mean	Std. Dev.	Min	Max
Menu:	Deductible (current NIS)[a]	Low	105,800	873.43	119.81	374.92	1,039.11
		Regular	105,800	1,455.72	199.68	624.86	1,731.85
		High	105,800	2,620.30	359.43	1,124.75	3,117.33
		Very High	105,800	3,784.87	519.18	1,624.64	4,502.81
	Premium (current NIS)[a]	Low	105,800	3,380.57	914.04	1,324.71	19,239.62
		Regular	105,800	3,189.22	862.30	1,249.72	18,150.58
		High	105,800	2,790.57	754.51	1,093.51	15,881.76
		Very High	105,800	2,551.37	689.84	999.78	14,520.46
	$\Delta p/\Delta d$		105,800	0.328	0.06	0.30	1.80
Realization:	Choice	Low	105,800	0.178	0.38	0	1
		Regular	105,800	0.811	0.39	0	1
		High	105,800	0.006	0.08	0	1
		Very High	105,800	0.005	0.07	0	1
	Policy Termination	Active	105,800	0.150	0.36	0	1
		Canceled	105,800	0.143	0.35	0	1
		Expired	105,800	0.707	0.46	0	1
	Policy Length (years)		105,800	0.848	0.28	0.005	1.08
	Claims	All	105,800	0.208	0.48	0	5
		Low	18,799	0.280	0.55	0	5
		Regular	85,840	0.194	0.46	0	5
		High	654	0.109	0.34	0	3
		Very High	507	0.107	0.32	0	2
	Claims per year[b]	All	105,800	0.245	0.66	0.00	198.82
		Low	18,799	0.309	0.66	0.00	92.64
		Regular	85,840	0.232	0.66	0.00	198.82
		High	654	0.128	0.62	0.00	126.36
		Very High	507	0.133	0.50	0.00	33.26

[a] The average exchange rate throughout the sample period was approximately 1 US dollar per 3.5 NIS, starting at 1:3 in late 1994 and reaching 1:4 in late 1999.

[b] The mean and standard deviation of the claims per year are weighted by the policy length to adjust for variation in the exposure period. These are the Maximum Likelihood estimates of a simple Poisson model with no covariates.

Table 2B: Contract Choices and Realizations

Claims	"Low"	"Regular"	"High"	"Very High"	Total	Share
0	11,929 (19.3%)	49,281 (79.6%)	412 (0.7%)	299 (0.5%)	61,921 (100%)	80.343%
1	3,124 (23.9%)	9,867 (75.5%)	47 (0.4%)	35 (0.3%)	13,073 (100%)	16.962%
2	565 (30.8%)	1,261 (68.8%)	4 (0.2%)	2 (0.1%)	1,832 (100%)	2.377%
3	71 (31.4%)	154 (68.1%)	1 (0.4%)	–	226 (100%)	0.293%
4	6 (35.3%)	11 (64.7%)	–	–	17 (100%)	0.022%
5	1 (50.0%)	1 (50.0%)	–	–	2 (100%)	0.003%

The table presents tabulation of choices and number of claims. To make things comparable, the figures are computed only for individuals whose policies lasted at least 0.9 years (about 73% of the data). The bottom rows of Table 2A provide descriptive figures for the full data. The percentages in parentheses present the distribution of deductible choices, conditional on the number of claims. The right-hand-side column presents the marginal distribution of the number of claims.

Table 3: Poisson Regressions (Dependent Variable: Number of Claims)

	Variable		IRR[a]	z-stat	p-value
Demographics:	Age		0.992	-1.49	0.137
	Age^2		1.0001	1.92	0.055
	Female		0.955	-2.83	0.005
	Family	Single	0.873	-0.77	0.441
		Married	0.782	-1.41	0.158
		Divorced	0.939	-0.36	0.720
		Widower	0.887	-0.66	0.508
	Education	Elementary	0.939	-1.16	0.247
		High School	0.989	-0.62	0.535
		Technical	1.026	0.85	0.396
		Academic	0.917	-4.61	0.000
	Emigrant		1.021	1.28	0.200
Car Attributes:	Log(Value)		1.127	4.28	0.000
	Car Age		1.018	4.30	0.003
	Commercial Car		0.869	-4.32	0.000
	Log(Engine)		1.349	6.53	0.000
Driving:	License Years		0.980	-5.95	0.000
	License $Years^2$		1.0002	3.24	0.001
	Good Driver		0.983	-0.9	0.367
	"Any Driver"		0.945	-3.33	0.001
	Secondary Car		0.918	-4.10	0.000
	Business Use		1.204	6.33	0.000
	History		0.949	-4.76	0.000
	Claim History		1.930	16.87	0.000
Young Driver:	Age	17-19	dropped		
		19-21	1.072	1.31	0.191
		21-24	0.973	-0.48	0.633
		>24	0.811	-3.69	0.000
	Experience	<1	dropped		
		1-3	0.785	-5.27	0.000
		>3	0.754	-5.31	0.000
	Gender	Male	1.689	14.54	0.000
		Female	1.457	9.39	0.000
Company Year:	First year		dropped		
	Second year		0.915	-4.49	0.000
	Third year		0.933	-3.12	0.002
	Fourth year		0.834	-7.83	0.000
	Fifth year		0.581	-19.38	0.000
Obs			105,798		
Pseudo R^2			0.016		

[a] IRR = Incidence Rate Ratio. Each figure should be interpreted as the increase/decrease in claim probability as a result of an increase of one unit in the right-hand-side variable.

Variation in exposure (policy length) is accounted for.

Table 4A: Probit Regressions (Dependent Variable: 1 if Low Deductible Chosen)

Variable			(1)		(2)		(3)	
			dP/dX	$z-stat$	dP/dX	$z-stat$	dP/dX	$z-stat$
Menu:	$\Delta p/\Delta d$		-		-0.354	-13.83	-0.351	-13.74
	$\overline{d}$		-		0.00016	14.96	0.00016	14.81
	$\widehat{\lambda}$		-		-		-0.153	-2.53
Demographics:	Age		-0.004	-4.85	-0.004	-4.73	-0.004	-5.13
	Age^2		$4.5*10^{-5}$	5.13	$4.4*10^{-5}$	5.06	$4.9*10^{-5}$	5.51
	Female		0.013	5.09	0.014	5.29	0.012	4.41
	Family	Single	0.044	1.24	0.038	1.1	0.032	0.92
		Married	0.043	1.38	0.038	1.23	0.028	0.90
		Divorced	0.050	1.37	0.046	1.28	0.042	1.18
		Widower	0.042	1.14	0.036	1.01	0.030	0.84
	Education	Elementary	-0.0010	-0.11	0.0005	0.06	-0.0019	-0.22
		High School	-0.0025	-0.83	-0.0010	-0.33	-0.0013	-0.44
		Technical	0.0111	2.24	0.0127	2.56	0.0139	2.79
		Academic	0.0027	0.90	0.0049	1.62	0.0015	0.45
	Emigrant		$2.2*10^{-4}$	0.08	$-9.3*10^{-6}$	0.00	$7.8*10^{-4}$	0.30
Car Attributes:	Log(Value)		0.030	6.48	0.030	5.64	0.035	6.16
	Car Age		$-1.7*10^{-3}$	-2.55	$3.3*10^{-5}$	0.05	$7.4*10^{-4}$	1.02
	Commercial Car		-0.029	-5.78	-0.027	-5.38	-0.032	-5.93
	Log(Engine)		0.008	1.02	0.003	0.42	0.015	1.71
Driving:	License Years		$5.1*10^{-4}$	0.89	$7.9*10^{-4}$	1.38	$-5.6*10^{-5}$	-0.08
	License Years^2		$-1.6*10^{-5}$	-1.46	$-2.0*10^{-5}$	-1.8	$-1.1*10^{-5}$	-0.9
	Good Driver		-0.015	-5.03	-0.012	-3.81	-0.012	-3.92
	"Any Driver"		-0.026	-9.94	-0.024	-9.23	-0.026	-9.57
	Secondary Car		-0.007	-2.17	-0.005	-1.52	-0.008	-2.26
	Business Use		-0.002	-0.32	-0.002	-0.37	0.006	1.01
	History		0.017	8.17	0.018	8.42	0.015	6.10
	Claim History		0.050	6.79	0.046	6.32	0.035	5.31
Young Driver:	Age	17-19	dropped		dropped		dropped	
		19-21	-0.015	-1.51	-0.014	-1.44	-0.010	-1.04
		21-24	-0.016	-1.50	-0.013	-1.23	-0.014	-1.30
		>24	0.013	1.26	0.014	1.30	0.004	0.36
	Experience	<1	dropped		dropped		dropped	
		1-3	-0.001	-0.11	-0.002	-0.20	-0.015	-1.49
		>3	0.041	3.85	0.037	3.48	0.019	1.62
	Gender	Male	-0.001	-0.15	-0.001	-0.11	0.027	2.03
		Female	0.018	2.33	0.018	2.28	0.038	3.40
Company Year:	First year		dropped		dropped		dropped	
	Second year		-0.086	-33.35	-0.088	-34.08	-0.091	-31.64
	Third year		-0.137	-50.57	-0.138	-51.65	-0.140	-49.85
	Fourth year		-0.173	-65.16	-0.173	-65.43	-0.176	-54.51
	Fifth year		-0.208	-72.88	-0.207	-72.66	-0.213	-40.85
Obs			105,798		105,798		105,798	
Pseudo R^2			0.1296		0.1343		0.1344	
log(Likelihood)			-43,085		-42,848		-42,845	

Table 4B: Probit Regressions (Dependent Variable: 1 if Low Deductible Chosen)

	Variable		(4)		(5)		(6)	
			dP/dX	$z-stat$	dP/dX	$z-stat$	dP/dX	$z-stat$
Menu:	$\Delta p/\Delta d$		-0.345	-12.92	-0.354	-13.83	-	
	$\overline{d}$		0.00014	13.14	0.00014	12.83	-	
	$\widehat{\lambda}$		-0.199	-1.28	-0.210	-1.29	-	
	$-log\left(\frac{\Delta p/(\widehat{\lambda}\Delta d)-1}{\overline{d}}\right)$		-		0.0006	0.23	0.0208	11.52
Demographics:	Age		-0.004	-4.46	-0.004	-4.46	-0.004	-3.57
	Age^2		4.7*10^{-5}	4.69	4.7*10^{-5}	4.69	3.3*10^{-5}	3.51
	Female		0.010	3.31	0.010	3.31	0.016	6.04
	Family	Single	0.031	0.73	0.031	0.73	0.068	1.51
		Married	0.025	0.63	0.025	0.63	0.066	1.84
		Divorced	0.037	0.86	0.037	0.86	0.061	1.34
		Widower	0.025	0.58	0.025	0.58	0.061	1.31
	Education	Elementary	-0.0038	-0.42	-0.0038	-0.43	0.0041	0.47
		High School	-0.0018	-0.58	-0.0018	-0.58	-0.0012	-0.38
		Technical	0.0138	2.59	0.0138	2.59	0.0087	1.68
		Academic	-0.0008	-0.19	-0.0008	-0.19	0.0096	3.05
	Emigrant		0.0014	0.51	0.0014	0.51	-0.0011	-0.41
Car Attributes:	Log(Value)		0.035	4.91	0.035	4.91	0.027	5.68
	Car Age		0.0006	0.63	0.0006	0.63	-0.0032	-4.57
	Commercial Car		-0.033	-4.78	-0.033	-4.78	-0.017	-3.16
	Log(Engine)		0.019	1.42	0.019	1.43	-0.014	-1.78
Driving:	License Years		-0.0006	-0.57	-0.0006	-0.57	0.0021	3.31
	License Years^2		-1.7*10^{-6}	-0.12	-1.7*10^{-6}	-0.12	-3.0*10^{-5}	-2.53
	Good Driver		-0.013	-3.78	-0.013	-3.79	-0.015	-4.57
	"Any Driver"		-0.027	-8.27	-0.027	-8.27	-0.022	-8.23
	Secondary Car		-0.008	-1.95	-0.008	-1.95	0.0002	0.06
	Business Use		0.009	1.05	0.009	1.05	-0.016	-3.00
	History		0.015	4.17	0.015	4.17	0.023	7.84
	Claim History		0.090	3.38	0.090	3.37	-0.010	-0.92
Young Driver:	Age	17-19	-0.007	-0.41	-0.007	-0.41	-0.035	-2.58
		19-21	-0.010	-0.72	-0.010	-0.72	-0.046	-4.86
		21-24	-0.018	-1.49	-0.018	-1.49	-0.043	-4.39
		>24	dropped		dropped		dropped	
	Experience	<1	-0.017	-1.01	-0.017	-1.01	-0.056	-4.43
		1-3	-0.033	-4.29	-0.033	-4.29	-0.041	-5.35
		>3	dropped		dropped		dropped	
	Gender	Male	0.053	4.32	0.053	4.32	0.049	4.00
		Female	0.068	4.85	0.068	4.84	0.095	6.82
Company Year:	First year		dropped		dropped		dropped	
	Second year		-0.088	-22.74	-0.088	-22.74	-0.079	-29.57
	Third year		-0.135	-41.05	-0.135	-41.03	-0.130	-47.14
	Fourth year		-0.171	-33.99	-0.171	-33.98	-0.162	-56.23
	Fifth year		-0.215	-19.15	-0.216	-19.14	-0.192	-51.53
Obs			93,988		93,988		93,988	
Pseudo R^2			0.1399		0.1399		0.1365	
log(Likelihood)			-36,808		-36,808		-36,954	

Table 4C: "Structural Interpretation" of the Probit Regressions

	Variable		Coef.	Std. Err.	$z-stat$
	Constant		-24.05	3.04	-7.92
	$-log\left(\frac{\Delta p/(\widehat{\lambda}\Delta d)-1}{\overline{d}}\right)$		One	-	-
Demographics:	Age		-0.16	0.05	-3.57
	Age^2		0.0016	0.0005	3.51
	Female		0.78	0.13	6.04
	Family	Single	2.94	1.95	1.51
		Married	3.57	1.94	1.84
		Divorced	2.62	1.95	1.34
		Widower	2.59	1.98	1.31
	Education	Elementary	0.20	0.42	0.47
		High School	-0.06	0.15	-0.38
		Technical	0.41	0.25	1.68
		Academic	0.46	0.15	3.05
	Emigrant		-0.05	0.13	-0.41
Car Attributes:	Log(Value)		1.30	0.23	5.68
	Car Age		-0.15	0.03	-4.57
	Commercial Car		-0.86	0.27	-3.16
	Log(Engine)		-0.69	0.39	-1.78
Driving:	License Years		0.101	0.03	3.31
	License Years^2		-0.0014	0.0006	-2.53
	Good Driver		-0.71	0.15	-4.57
	"Any Driver"		-1.07	0.13	-8.23
	Secondary Car		0.01	0.16	0.06
	Business Use		-0.80	0.27	-3.00
	History		1.12	0.14	7.84
	Claim History		-0.49	0.53	-0.92
Young Driver:	Age	17-19	-1.86	0.72	-2.58
		19-21	-2.54	0.52	-4.86
		21-24	-2.32	0.53	-4.39
		>24	dropped		
	Experience	<1	-3.21	0.72	-4.43
		1-3	-2.18	0.41	-5.35
		>3	dropped		
	Gender	Male	2.12	0.53	4.00
		Female	3.86	0.57	6.82
Year Dummies:			yes		
σ			10.35		
Obs			93,988		
Pseudo R^2			0.1365		
log(Likelihood)			-36,954		

This regression is a replication of column (6) from Table 4B. It presents a structural interpretation of the results by reporting coefficients (not changes in probabilities) and renormalizing the coefficients by the coefficient on the cutoff point (freeing up the variance of the error term). This, together with the assumption that the coefficient of absolute risk aversion, r, follows a Lognormal distribution, allows us to interpret the coefficients as if it is a linear regression in which the dependent variable is $log(r)$. One should be cautious in interpreting these coefficients, however. Unlike the full structural model, this regression does not allow unobserved heterogeneity in risk and suffers from some selection bias because observations with "too high" predicted risk rate are omitted. Thus, it is only useful for comparison.

Table 5: The Benchmark Model

Variable			Dep Var: $\log(\lambda)$ Coef.	Std. Err.	Dep Var: $\log(r)$ Coef.	Std. Err.
Demographics:	Constant		-1.572	0.007	-11.629	0.096
	Age		-0.0001	0.0006	0.0035	0.005
	Female		0.004	0.008	0.161	0.061
	Family	Single	0.0063	0.100	0.513	0.778
		Married	0.0493	0.099	0.598	0.773
		Divorced	0.1030	0.099	0.231	0.779
		Widower	0.0740	0.102	0.394	0.791
	Education	Elementary	-0.069	0.028	0.406	0.192
		High School	-0.053	0.011	0.313	0.080
		Technical	-0.063	0.017	0.570	0.114
		Academic	-0.083	0.012	0.525	0.078
	Emigrant		-0.0023	0.0094	0.0197	0.0661
Car Attributes:	Log(Value)		0.085	0.017	0.735	0.117
	Car Age		-0.0022	0.0022	0.0019	0.0163
	Commercial Car		-0.074	0.018	-0.002	0.120
	Log(Engine)		0.172	0.023	-0.532	0.180
Driving:	License Years		-0.0019	0.0007	0.005	0.005
	Good Driver		-0.058	0.010	-0.137	0.072
	"Any Driver"		-0.055	0.0097	-0.197	0.067
	Secondary Car		-0.034	0.014	0.075	0.087
	Business Use		0.050	0.014	-0.331	0.110
	History		-0.002	0.005	0.296	0.046
	Claim History		0.144	0.016	-0.060	0.159
Young Driver:	Age	17-19	0.054	0.016	-	
		19-21	-0.035	0.012	-	
		21-24	-0.031	0.013	-	
		>24	0.032	0.011	-	
	Experience	<1	-0.004	0.011	-	
		1-3	0.089	0.012	-	
		>3	dropped		-	
	Gender	Male	0.038	0.006	-	
		Female	dropped		-	
Company Year:	First year		dropped		dropped	
	Second year		-0.342	0.010	-0.441	0.177
	Third year		-0.222	0.014	-1.987	0.129
	Fourth year		-0.283	0.015	-2.983	0.148
	Fifth year		-0.540	0.024	-3.028	0.150
σ			0.172	0.009	2.986	0.062
ρ			0.861	0.026		
Obs			105,798			

Table 6: CARA utility

Variable			Dep Var: log(λ) Coef.	Std. Err.	Dep Var: log(r) Coef.	Std. Err.
Demographics:	Constant		-1.583	0.007	-10.286	0.065
	Age		-0.0006	0.0006	0.0053	0.0032
	Female		0.009	0.009	0.067	0.051
	Family	Single	0.030	0.081	0.439	0.613
		Married	0.012	0.080	0.548	0.611
		Divorced	0.064	0.081	0.262	0.621
		Widower	0.040	0.084	0.379	0.630
	Education	Elementary	-0.061	0.031	0.317	0.143
		High School	-0.034	0.010	0.196	0.056
		Technical	-0.035	0.017	0.323	0.090
		Academic	-0.063	0.011	0.354	0.059
	Emigrant		0.0012	0.0079	0.0003	0.0435
Car Attributes:	Log(Value)		0.043	0.015	0.726	0.098
	Car Age		-0.0017	0.0020	0.0069	0.0116
	Commercial Car		-0.086	0.017	0.162	0.083
	Log(Engine)		0.172	0.021	-0.523	0.133
Driving:	License Years		-0.0015	0.0007	0.0031	0.0037
	Good Driver		-0.057	0.009	-0.046	0.049
	"Any Driver"		-0.053	0.009	-0.070	0.049
	Secondary Car		-0.027	0.012	0.066	0.061
	Business Use		0.053	0.014	-0.275	0.080
	History		0.002	0.005	0.178	0.038
	Claim History		0.155	0.017	-0.264	0.118
Young Driver:	Age	17-19	0.067	0.017	-	
		19-21	-0.023	0.012	-	
		21-24	-0.024	0.013	-	
		>24	0.028	0.012	-	
	Experience	<1	-0.005	0.011	-	
		1-3	0.071	0.012	-	
		>3	dropped		-	
	Gender	Male	0.030	0.006	-	
		Female	dropped		-	
Company Year:	First year		dropped		dropped	
	Second year		-0.249	0.007	-0.436	0.147
	Third year		-0.224	0.014	-1.024	0.086
	Fourth year		-0.289	0.016	-1.563	0.098
	Fifth year		-0.515	0.023	-1.345	0.104
σ			0.211	0.008	1.522	0.031
ρ			0.852	0.021		
Obs			105,798			

This regression is a replication of Table 5 for a CARA utility specification, i.e. the deductible choice is given by equation (8).

Table 7: Implications for risk aversion levels

Specification[a]	ARA^b	Interpretation[c]	RRA^d
Benchmark (mean)	$1.6 \cdot 10^{-3}$	61.27	81.87
Benchmark (median)	$7.8 \cdot 10^{-6}$	99.72	0.40
CARA (mean)	$1.9 \cdot 10^{-4}$	93.76	9.67
CARA (median)	$3.0 \cdot 10^{-5}$	98.96	1.52
Incomplete Info (mean)	$1.1 \cdot 10^{-3}$	70.91	56.28
Incomplete Info (median)	$1.5 \cdot 10^{-6}$	99.95	0.08
Gertner (1993)	$3.1 \cdot 10^{-4}$	96.99	4.79
Metrick (1995)	$6.6 \cdot 10^{-5}$	99.34	1.02
Holt and Laury (2002)[e]	$3.2 \cdot 10^{-2}$	20.96	865.75

[a] This table summarizes the results with respect to the level of risk aversion. "Benchmark" refers to the results from the benchmark model (Table 5), "CARA" refers to a specification of a CARA utility function (Table 6), and "Incomplete Info" refers to a specification in which individuals do not know their risk types perfectly (Table 9). The last three rows are the closest comparable results available in the literature.

[b] The second column presents the point estimates for the coefficient of absolute risk aversion, converted to $\$US^{-1}$ units. For Gertner (1993), Metrick (1995), and Holt and Laury (2002) this is given by their estimate of a representative CARA utility maximizer. For all other specifications, this is given by computing the unconditional mean and median in the population. This accounts for the variation in observables as well. Since we use Gibbs sampler, we augment the latent individual risk aversion levels into the estimation, so computing the mean and median from the posterior distribution is straightforward.

[c] To interpret the ARA estimates, we translate them into the following quantity $\{x : u(w) = \frac{1}{2}u(w+100) + \frac{1}{2}u(w-x)\}$. Namely, we report x such that an individual with the estimated ARA is indifferent about participating in a fifty-fifty lottery of gaining 100 \$US and losing x \$US. Note that since our estimate is of absolute risk aversion, the quantity x is independent of w. To be consistent with the specification, we use a quadratic utility function for the benchmark and incomplete information cases, and use a CARA utility function for the others.

[d] The last column attempts to translate the ARA estimates into relative risk aversion. We follow the literature, and do so by multiplying the ARA estimate by average annual income. For this, we use the average annual income (after tax) in Israel in 1995 (from Israeli census) for all our specifications, and we use average disposable income in the US in 1987 for Gertner (1993) and Metrick (1995), which is the same number used by Gertner (1993). For Holt and Laury (2002) we use a similar figure for 2002.

[e] Holt and Laury (2002) do not report a comparable estimate. The estimate we provide above is based on estimating a CARA utility model for the 18 subjects in their experiment who participated in the "$\times 90$" treatment, which involved stakes comparable to our setting. For these subjects, we assume a CARA utility function and a Lognormal distribution of their coefficient of absolute risk aversion. The table reports the point estimate of the mean from this distribution.

Table 8: Experienced Drivers

Variable			Dep Var: log(λ)		Dep Var: log(r)	
			Coef.	Std. Err.	Coef.	Std. Err.
Demographics:	Constant		-1.597	0.008	-11.441	0.100
	Age		0.0001	0.0007	0.0095	0.0051
	Female		0.006	0.011	0.102	0.078
	Family	Single	0.048	0.107	0.507	0.821
		Married	0.045	0.106	0.532	0.813
		Divorced	0.106	0.107	0.151	0.826
		Widower	0.067	0.111	0.304	0.848
	Education	Elementary	-0.098	0.035	0.519	0.206
		High School	-0.043	0.012	0.256	0.085
		Technical	-0.054	0.018	0.533	0.127
		Academic	-0.090	0.013	0.550	0.083
	Emigrant		-0.0059	0.0096	0.052	0.067
Car Attributes:	Log(Value)		0.070	0.021	0.801	0.139
	Car Age		-0.0029	0.0027	-0.0001	0.018
	Commercial Car		-0.090	0.021	0.098	0.129
	Log(Engine)		0.186	0.026	-0.681	0.197
Driving:	License Years		-0.002	0.0009	-0.002	0.006
	Good Driver		-0.054	0.011	-0.161	0.076
	"Any Driver"		-0.055	0.011	-0.191	0.073
	Secondary Car		-0.039	0.014	0.112	0.087
	Business Use		0.050	0.015	-0.347	0.115
	History		-0.010	0.007	0.314	0.064
	Claim History		0.195	0.021	-0.325	0.187
Young Driver:	Age	17-19	0.070	0.017	-	
		19-21	-0.038	0.016	-	
		21-24	-0.010	0.017	-	
		>24	0.023	0.016	-	
	Experience	<1	0.012	0.014	-	
		1-3	0.072	0.017	-	
		>3	dropped		-	
	Gender	Male	0.041	0.008	-	
		Female	dropped		-	
Company Year:	First year		dropped		dropped	
	Second year		-0.352	0.012	-0.526	0.197
	Third year		-0.241	0.015	-1.920	0.126
	Fourth year		-0.303	0.017	-2.941	0.153
	Fifth year		-0.551	0.026	-2.967	0.155
σ			0.188	0.009	2.930	0.070
ρ			0.784	0.037		
Obs			82,964			

This regression is a replication of Table 5 for only experienced drivers, i.e. drivers who had driving license for ten years or more.

Table 9: Incomplete Information of Risk Type

Variable			Dep Var: log(λ)		Dep Var: log(r)	
			Coef.	Std. Err.	Coef.	Std. Err.
Demographics:	Constant		-1.680	0.006	-13.441	0.180
	Age		0.0003	0.0008	0.013	0.007
	Female		0.004	0.011	0.181	0.103
	Family	Single	0.005	0.110	1.550	1.715
		Married	-0.068	0.109	1.592	1.715
		Divorced	0.012	0.109	1.526	1.713
		Widower	-0.010	0.111	1.542	1.715
	Education	Elementary	-0.089	0.038	0.713	0.294
		High School	-0.032	0.012	0.345	0.120
		Technical	-0.030	0.022	0.662	0.193
		Academic	-0.077	0.014	0.670	0.126
	Emigrant		0.0114	0.0106	-0.035	0.101
Car Attributes:	Log(Value)		0.131	0.018	1.924	0.201
	Car Age		0.002	0.0026	-0.019	0.025
	Commercial Car		-0.110	0.022	0.283	0.177
	Log(Engine)		0.241	0.027	-1.125	0.286
Driving:	License Years		-0.0034	0.0009	0.010	0.008
	Good Driver		-0.0778	0.0112	-0.237	0.107
	"Any Driver"		-0.086	0.011	-0.230	0.103
	Secondary Car		-0.039	0.014	0.112	0.129
	Business Use		0.076	0.018	-0.679	0.180
	History		-0.018	0.007	0.424	0.086
	Claim History		0.336	0.024	-0.824	0.290
Young Driver:	Age	17-19	0.148	0.021	-	
		19-21	0.044	0.023	-	
		21-24	0.045	0.024	-	
		>24	0.108	0.023	-	
	Experience	<1	-0.046	0.020	-	
		1-3	0.065	0.023	-	
		>3	dropped		-	
	Gender	Male	0.028	0.011	-	
		Female	dropped		-	
Company Year:	First year		dropped		dropped	
	Second year		-0.410	0.010	-1.586	0.423
	Third year		-0.426	0.013	-2.156	0.164
	Fourth year		-0.569	0.015	-3.400	0.209
	Fifth year		-0.883	0.019	-3.333	0.208
σ			0.279	0.010	3.377	0.105
ρ			0.723	0.042		
Obs			105,798			

This regression estimates a specification of the model where individuals do not have complete information about their risk types. Individuals are Bayesian and update their information from past claims history. We discuss this specification in the end of Section 4. The end of Appendix A provides the technical details.

Table 10: Representativeness

Variable		Sample[b]	Population[c]	Car Owners[d]
Age[a]		41.14 (12.37)	42.55 (18.01)	45.11 (14.13)
Female		0.316	0.518	0.367
Family	Single	0.143	0.233	0.067
	Married	0.780	0.651	0.838
	Divorced	0.057	0.043	0.043
	Widower	0.020	0.074	0.052
Education	Elementary	0.029	0.329	0.266
	High School	0.433	0.384	0.334
	Technical	0.100	0.131	0.165
	Academic	0.438	0.155	0.234
Emigrant		0.335	0.445	0.447
Obs		105,800	723,615	255,435

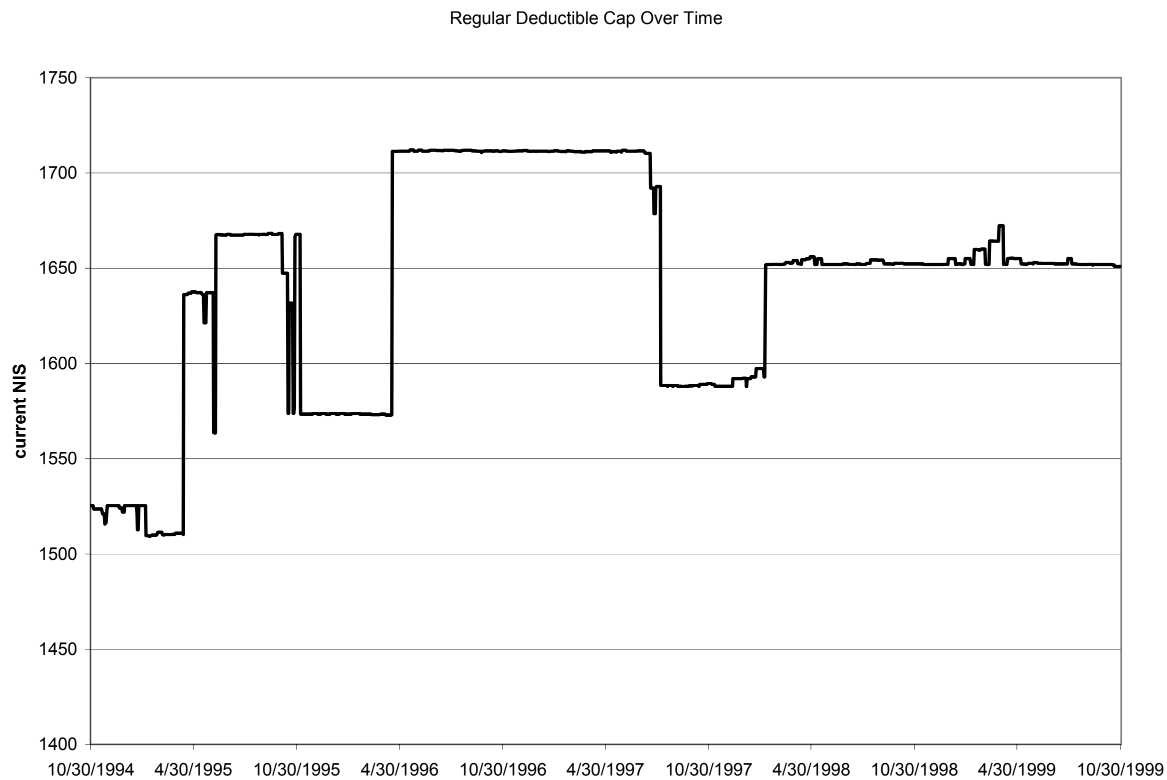
[a] For the age variable, the only continuous variable in the table, we provide both the mean and the standard deviation (in parentheses).

[b] The figures are identical to those presented in Table 1. The family and education variables are renormalized so they add up to 1 (i.e. we ignore those individuals for which we do not have family status or education level). This is particularly relevant for the education variable, as those who did not report it have probably not done so at random.

[c] This column is based on a random sample of the Israeli population as of 1995. We use only adult population, i.e. individuals who are 18 years old or more.

[d] This column is based on a subsample of the population sample. The data only provides information about car ownership at the household level, not at the individual level. Thus, according to our (rough) definition, an individual is a car owner if one of the following two conditions apply: (i) the household owns at least one car, and the individual is the head of the household; or (ii) the household owns at least two cars, and the individual is the spouse of the head of the household.

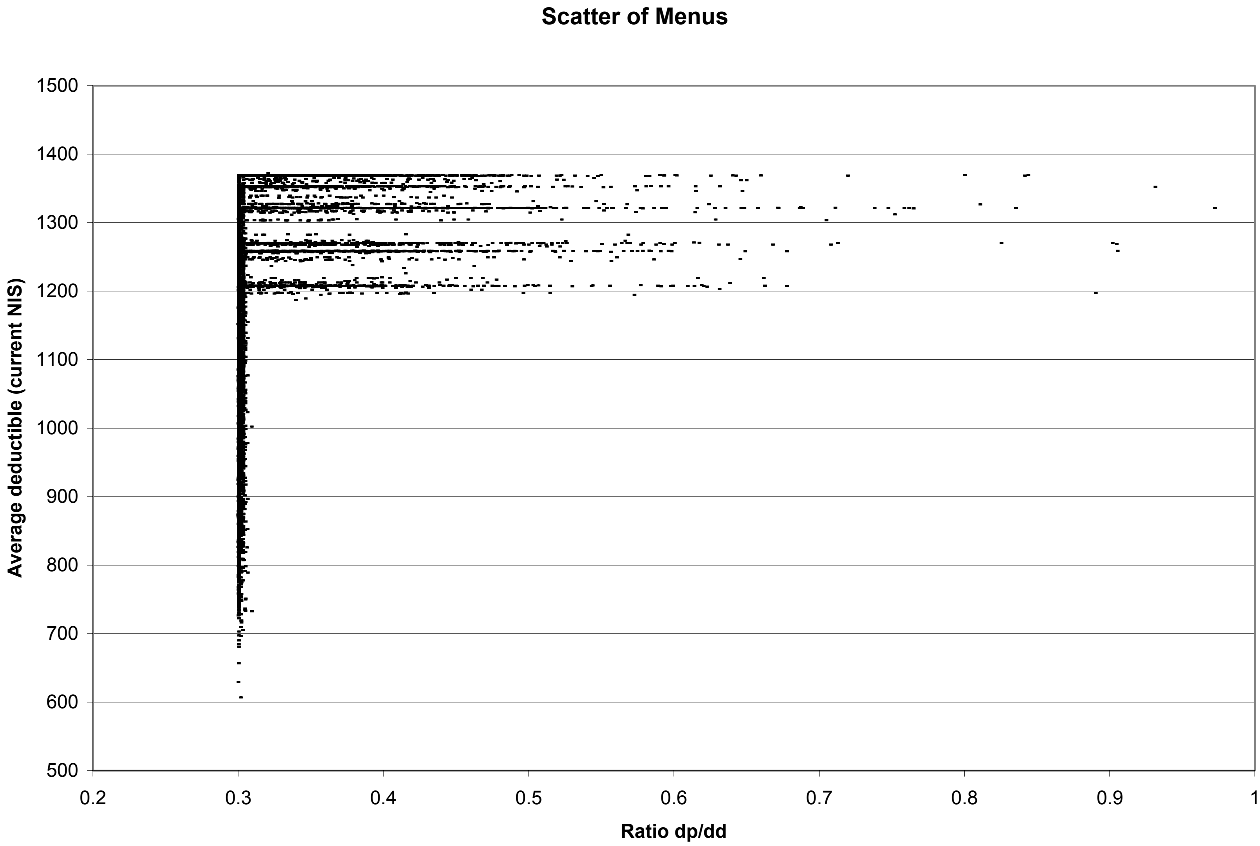
Figure 1: Variation in the Deductible Cap Over Time



This graph presents the variation in the deductible cap over time, which is the main source of (what we argue to be) exogenous variation in the data. While we do not observe the cap directly, the cap can be pretty accurately calculated from observing the menus offered. The graph above plots the maximal regular deductible offered to anyone who bought insurance from the company over a moving 7-day window. The big jumps in the graph reflect changes in the deductible cap.

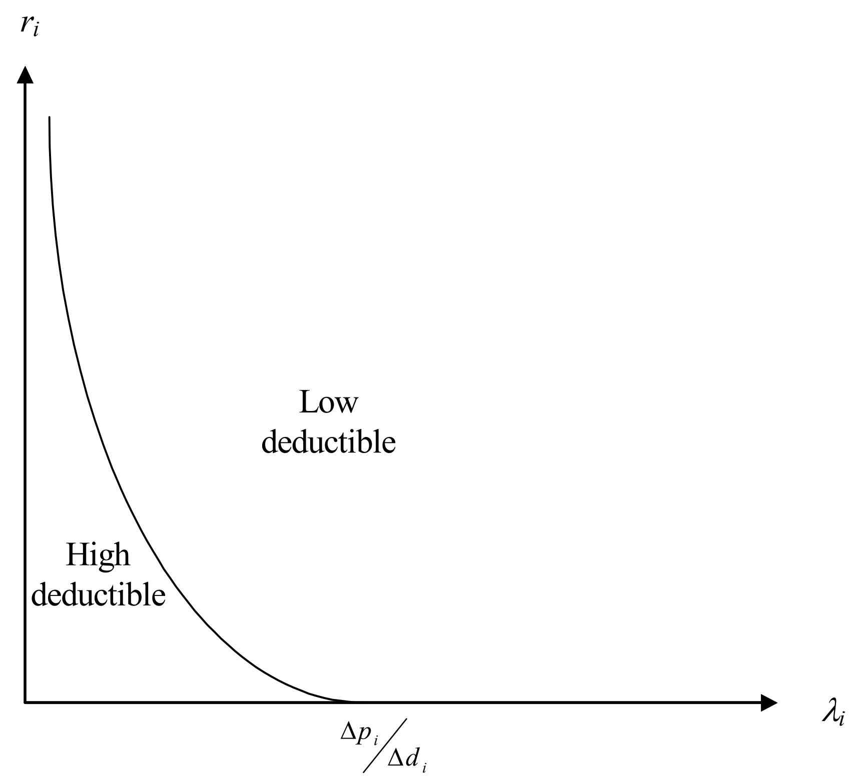
There are three reasons why the graph is not perfectly smooth. First, in few holiday periods (e.g. October 1995) there are not enough sales within a 7-day window, so none of those sales hits the cap. This gives rise to temporary jumps downwards. Second, the pricing rule applies at the date of the price quote given to the potential customer. Our recorded date is the first date the policy becomes effective. The price quote is held for a period of 2-4 weeks, so in periods in which the pricing changes, we may still see new policies which are sold using earlier quotes, made according to the previous pricing regime. Finally, even within periods of constant cap, the maximal deductible varies slightly (variation of less than 0.5%). We do not know the source of this variation.

Figure 2: Price (Menu) Variation



In the figure above we plot the two primary options of each menu offered in the data in the space $\frac{\Delta p_i}{\Delta d_i} = \frac{p_i^l - p_i^h}{d_i^h - d_i^l}$ and $\overline{d} = \frac{d_i^h + d_i^l}{2}$, which according to the model we present in Section 3.1 is the relevant space for the deductible choice. The goal is to provide some feel for the (unconditional) variation in menus we have in the data. The thick vertical line at 0.3 is driven by the pricing formula for individuals who do not hit the deductible cap. For individuals who hit the deductible cap (approximately one third of the data), the prices are higher, but the deductibles are fixed, so the ratio increases. One can see multiple "soft" horizontal lines. Each such line reflects a different level of the deductible cap, as shown in Figure 1.

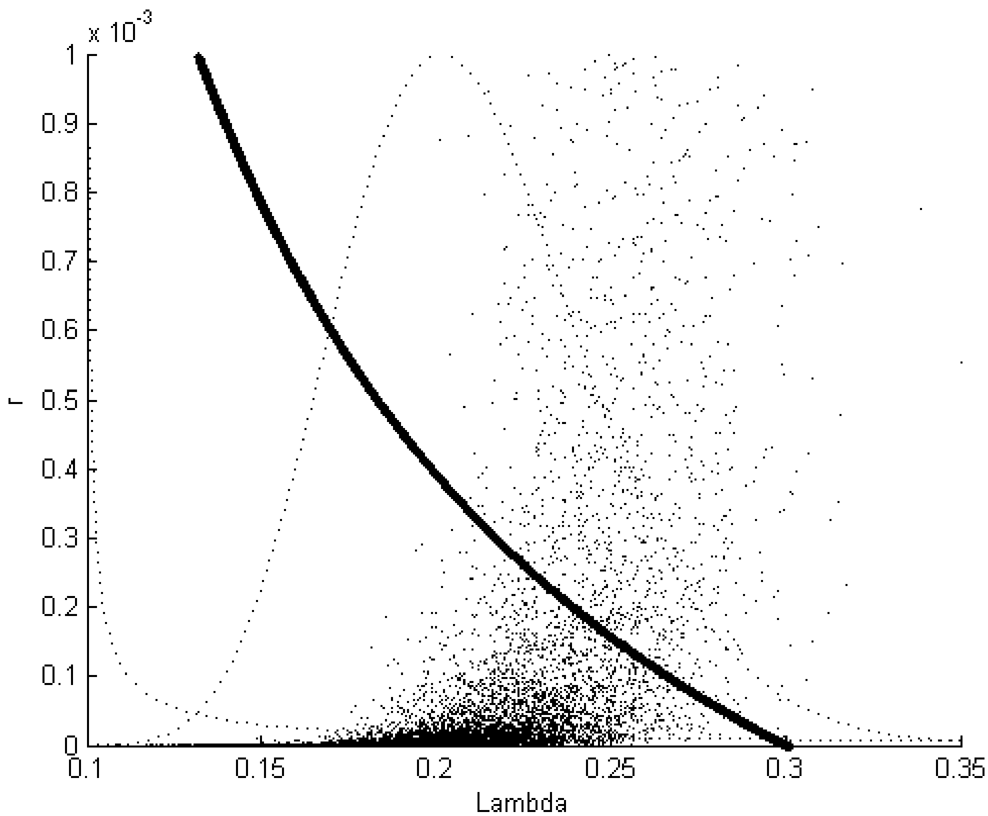
Figure 3: The Individual's Decision - A Graphical Illustration



This graph illustrates the individual's decision problem. An individual is a point in the above two-dimensional space. Each two deductible-premium combinations can be translated to an indifference set of points, given by the downwards sloping curve. If an individual is either to the right of the line (high risk) or above the line (high risk-aversion), the lower deductible would be optimal. Adverse selection is captured by the fact that the line is downward sloping: higher risk individuals require lower levels of risk aversion to make the low deductible choice. Thus, in the absence of correlation between risk and risk aversion, higher risk individuals are more likely to choose higher levels of insurance.

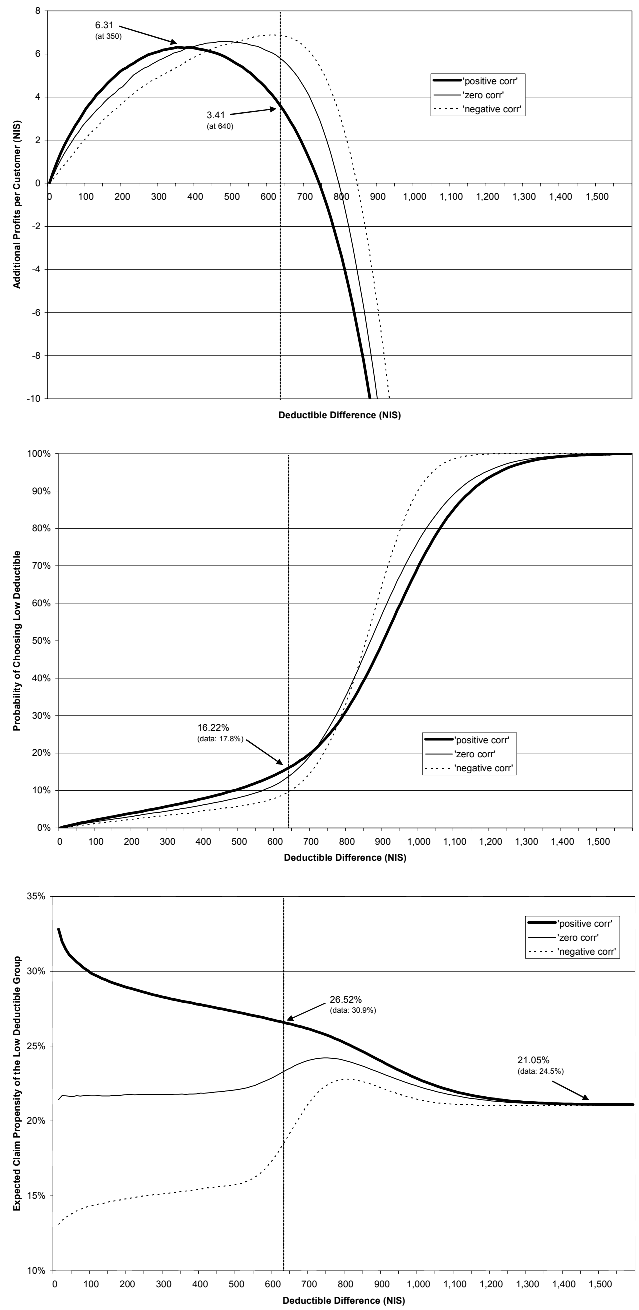
Two other important things to note about the graph. First, an individual with $\lambda_i > \frac{\Delta p_i}{\Delta d_i}$ will choose lower deductible even with risk-neutrality, i.e. with probability one (we do not allow individuals to be risk-loving). This does not create an estimation problem because λ_i is not observed, only a posterior distribution for it. Any such distribution will have a positive weight on values of λ_i which are below $\frac{\Delta p_i}{\Delta d_i}$. Second, the indifference set is a function of the menu, and in particular of $\frac{\Delta p_i}{\Delta d_i}$ and $\overline{d}$. An increase in $\frac{\Delta p_i}{\Delta d_i}$ will shift the set up and to the right, and an increase in $\overline{d}$ will shift the set down and to the left. Therefore, exogenous shifts of the menus that make both arguments change in the same direction can make the sets "cross," and thereby allowing to nonparametrically identify the correlation between risk and risk aversion.

Figure 4: Graphical illustration of the results



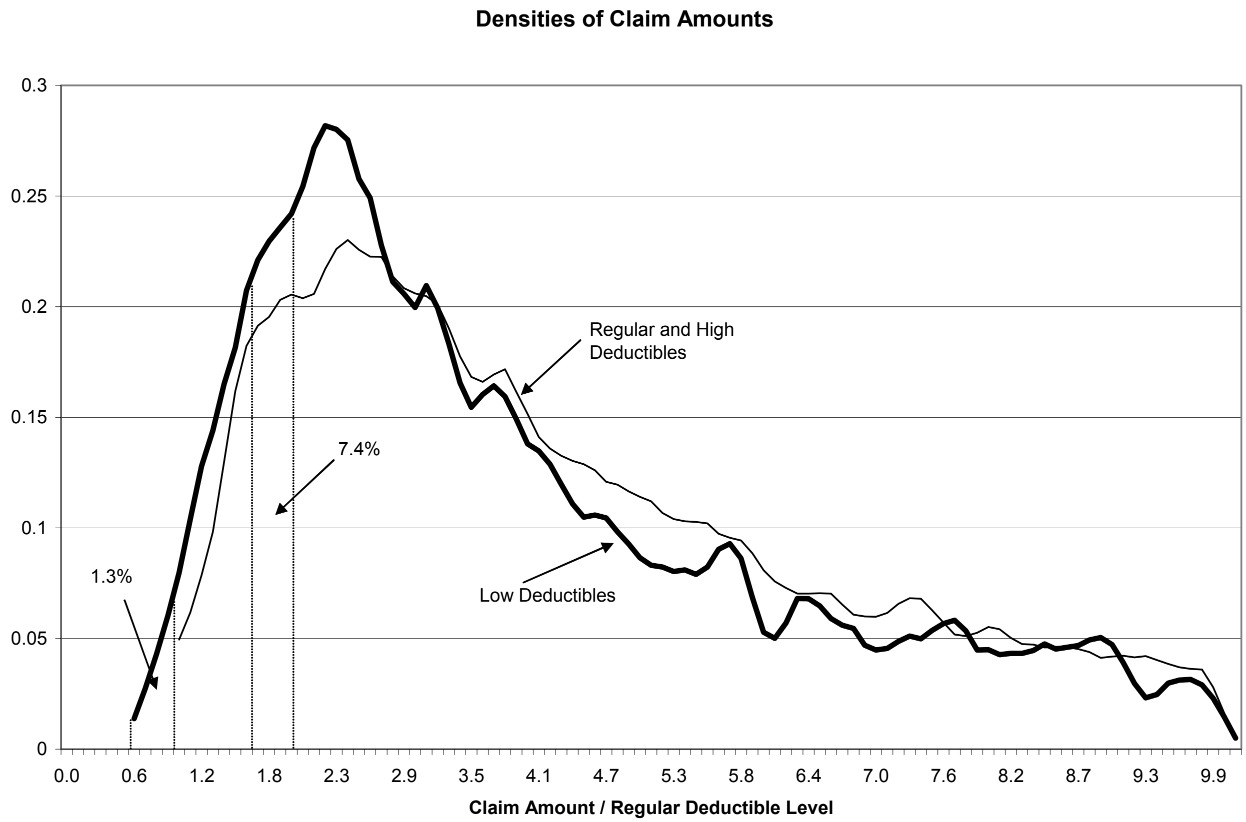
This figure illustrates the results for the benchmark model in the space of (λ_i, r_i). The solid line presents the indifference set (equation (7)) applied for the menu faced by the average individual in the sample. The two dotted lines present normalized marginal (lognormal) densities of λ_i and r_i for the average individual, based on the point estimates of the benchmark model (Table 5). The scattered points present 10,000 draws from the joint distribution. The figure also illustrates how changing the deductible-premium menu is affected by adverse selection. A higher price (or higher low deductible) will shift the solid line up and to the right, affecting some of the individuals who previously chose a low deductible. Due to the positive correlation, these marginal individuals are relatively high risk, therefore creating a stringer incentive for the insurer to raise the price of the low deductible.

Figure 5: Counterfactuals - Varying the level of the low deductible



This figure illustrates the results from the counterfactual exercise (see Section 4.3). On the horizontal axis, we change the deductible benefit of low deductible, $\Delta d = d_h - d_l$, in all graphs. The top panel presents the additional profits from offering the low deductible, the middle panel presents the fraction of consumers choosing low deductible, and the bottom panel presents the average risk of those individuals who choose low deductible. The vertical line in all graphs presents the actual level of low deductible, thereby providing some indication for the fit of the model. As the top panel shows, while offering the "low" alternative raises profits, these profits could be higher by making this alternative less attractive.

Figure 6: Claim Distributions



In the figure above we plot kernel densities of the claim amounts, normalized by the level of the regular deductible (there is an additional fat tail outside the figure, which accounts for about 25% of the distribution). The thick line presents the distribution of the claim amounts for individuals who chose low deductible, while the thin line does the same for those who chose regular deductible. Clearly, both distributions are truncated from below at the deductible level. The figure shows that the distributions are fairly similar, and that the probability of a claim falling above the low deductible but below the regular deductible is very small (1.3%). This implies that assuming that only claim rate matters for the deductible choice (but not the claim amount) is not very restrictive.