

ISSN 1045-6333

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

ESTIMATING RISK PREFERENCES FROM DEDUCTIBLE CHOICE

Alma Cohen and Liran Einav

Discussion Paper No. 582

03/2007

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

The Social Science Research Network Electronic Paper Collection:
http://papers.ssrn.com/abstract_id=977586

Estimating Risk Preferences from Deductible Choice*

By Alma Cohen[†] and Liran Einav[‡]

Abstract: *We estimate the distribution of risk preferences using a large data set of deductible choices in auto insurance contracts. To do so, we develop a structural econometric model of adverse selection that allows for unobserved heterogeneity in both risk (claim rate) and risk aversion. We use data on realized claims to estimate the distribution of claim rates and data on deductible and premium choices to estimate the distribution of risk aversion and how it correlates with risk. We find large heterogeneity in risk attitudes: while the majority of individuals are almost risk neutral with respect to lotteries of 100 dollar magnitude, an important fraction of the individuals exhibit significant risk aversion even with respect to such relatively small bets. The estimates imply that women are more risk averse than men, that risk aversion exhibits a U-shape with respect to age, and that most proxies for income and wealth are positively associated with absolute risk aversion. Finally, unobserved heterogeneity in risk aversion is more important than that of risk, and risk and risk aversion are positively correlated.*

Journal of Economic Literature classification numbers: D82, G22.

Keywords: risk aversion, adverse selection, structural estimation, mixture models.

*We thank two anonymous referees and Judy Chevalier, the editor, for many helpful comments, which greatly improved the paper. We are also grateful to Susan Athey, Lucian Bebchuk, Lanier Benkard, Tim Bresnahan, Raj Chetty, Ignacio Esponda, Amy Finkelstein, Igal Hendel, Mark Israel, Felix Kubler, Jon Levin, Aviv Nevo, Harry Paarsch, Daniele Paserman, Peter Rossi, Esteban Rossi-Hansberg, Morten Sorensen, Manuel Trajtenberg, Frank Wolak, and seminar participants at UC Berkeley, CEMFI, Chicago GSB, Duke, Haas, Harvard, Hoover, the Interdisciplinary Center in Herzliya, Minnesota, Northwestern, Penn State, Stanford, Tel Aviv, UW Madison, Wharton, SITE 2004, the Econometric Society 2005 and 2006 winter meetings, and the NBER IO 2005 winter meeting for many useful discussions and suggestions. All remaining errors are, of course, ours. Einav acknowledges financial support from the National Science Foundation (Grant No. SES-0452555) and from Stanford's Office of Technology Licensing, and the hospitality of the Hoover Institution. Address correspondence to Liran Einav, Department of Economics, Stanford University, Stanford, CA 94305-6072; Tel: (650) 723-3704; E-mail: leinav@stanford.edu.

[†]Department of Economics, Tel-Aviv University, National Bureau of Economic Research, and Harvard Law School John M. Olin Research Center for Law, Economics, and Business. E-mail: almac@post.tau.ac.il.

[‡]Department of Economics, Stanford University, and National Bureau of Economic Research. E-mail: leinav@stanford.edu.

1 Introduction

The analysis of decisions under uncertainty is central to many fields in economics, such as macroeconomics, finance, and insurance. In many of these applications it is important to know the degree of risk aversion, how heterogeneous individuals are in their attitudes towards risk, and how these attitudes vary with individuals' characteristics. Somewhat surprisingly, these questions have received only little attention in empirical microeconomics, so answering them using direct evidence from risky decisions made by actual market participants is important.

In this study, we address these questions by estimating risk preferences from the choice of deductible in insurance contracts. We use a rich data set of more than 100,000 individuals choosing from an individual-specific menu of deductible and premium combinations offered by an Israeli auto insurance company. An individual who chooses low deductible is exposed to less risk, but is faced with a higher level of expected expenditure. Thus, an individual's decision to choose a low (high) deductible provides a lower (upper) bound for his coefficient of absolute risk aversion.

Inferring risk preferences from insurance data is particularly appealing, as risk aversion is the primary reason for the existence of insurance markets. To the extent that extrapolating utility parameters from one market context to another necessitates additional assumptions, there is an advantage to obtaining such parameters from the same markets to which they are subsequently applied. The deductible choice is (almost) an ideal setting for estimating risk aversion in this context. Other insurance decisions, such as the choice among health plans, annuities, or just whether to insure or not, may involve additional preference-based explanations that are unrelated to financial risk and make inference about risk aversion difficult.¹ In contrast, the choice among different alternatives that vary only in their financial parameters (the levels of deductibles and premiums) is a case in which the effect of risk aversion can be more plausibly isolated and estimated.

The average deductible menu in our data offers an individual to pay an additional premium of 55 U.S. dollars in order to save 182 dollars in deductible payments in the event of a claim.² A risk-neutral individual should choose a low deductible if and only if his claim propensity is greater than the ratio between the premium (55) and the potential saving (182), which is 0.3. Although this pricing is actuarially unfair with respect to the average claim rate of 0.245, 18 percent of the sample choose to purchase it. Are these individuals exposed to greater risk than the average individual, are they more risk averse, or a combination of both? We answer this question by developing a structural econometric model and estimating the joint distribution of risk and risk aversion.

Our benchmark specification uses expected utility theory to model individuals' deductible choices as a function of two utility parameters, the coefficient of absolute risk aversion and a claim

¹For example, Matthew Rabin and Richard H. Thaler (2001, footnote 2) point out that one of their colleagues buys the insurance analyzed by Charles J. Cicchetti and Jeffrey A. 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 critique.

²For ease of comparison, we convert many of the reported figures from New Israeli Shekels to U.S. dollars. It is important to keep in mind, however, that GDP per capita in Israel was 0.52 – 0.56 of that in the U.S. (0.67 – 0.70, when adjusted for PPP) during the observation period.

rate. We allow both utility parameters to depend on individuals’ observable and unobservable characteristics, and assume that there is no moral hazard. Two key assumptions – that claims are generated by a Poisson process at the individual level, and that individuals have perfect information about their Poisson claim rates – allow us to use data on (ex post) realized claims to estimate the distribution of (ex ante) claim rates. Variation in the deductible menus across individuals and their choices from these menus are then used to estimate the distribution of risk aversion in the sample and the correlation between risk aversion and claim risk. Thus, we can estimate heterogeneous risk preferences from deductible choices, accounting for adverse selection (unobserved heterogeneity in claim risk), which is an important confounding factor.³

Our results suggest that heterogeneity in risk preferences is rather large. While the majority of the individuals are estimated to be close to risk neutral with respect to lotteries of 100 dollar magnitude, a significant fraction of the individuals in our sample exhibit significant levels of risk aversion even with respect to such relatively small bets. Overall, an individual with the average risk aversion parameter in our sample is indifferent about participating in a fifty-fifty lottery in which he gains 100 dollars or loses 56 dollars. We find that women are more risk averse than men, that risk preferences exhibit a U-shape with respect to age, and, interestingly, that most proxies for income and wealth are positively associated with absolute risk aversion.

We perform an array of tests to verify that these qualitative results are robust to deviation from the modeling assumptions. In particular, we explore alternative distributional assumptions, alternative restrictions on the von Neumann-Morgenstern (vNM) utility function, and a case in which individuals are allowed to make “mistakes” in their coverage choices due to incomplete information about their own risk types. We also show that the risk preferences we estimate are stable over time and help predict other (but closely related) insurance decisions. Finally, we justify our assumption to abstract from moral hazard, and we discuss the way this and other features of the setup (sample selection and additional costs associated with an accident) may affect the interpretation of the result.

Throughout we mostly focus on absolute (rather than relative) risk aversion.⁴ This allows us to take a neutral position with respect to the recent debate over the empirical relevance of expected utility theory (Matthew Rabin, 2000; Matthew Rabin and Richard H. Thaler, 2001; Ariel Rubinstein, 2001; Richard Watt, 2002; Nicholas Barberis et al., 2006; Ignacio Palacios-Huerta et al., 2006). While the debate focuses on how the curvature of the vNM utility function varies with wealth or across different settings, we only measure this curvature 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. Avoiding this debate is also a drawback. Without taking

³Throughout 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 refers to both these selection mechanisms as adverse selection, with the distinction being between *common values* (selection on risk) and *private values* (selection on risk aversion).

⁴Primarily as a way to compare our results with other estimates in the literature, Section 4 also provides estimates of relative risk aversion by following the literature and multiplying our estimates for absolute risk aversion by annual income (in Israel). We obtain high double-digit estimates for the mean individual, but below 0.5 for the median.

a stand on the way absolute risk preferences vary with sizes and contexts, we cannot discuss how relevant our estimates are for other settings. Obviously, we think they are. But since statements about their external relevance are mainly informed by what we think and less by what we do, we defer this discussion to the concluding section.

Our analysis also provides two results regarding the relationship between the distribution of risk preferences and that of risk. First, we find that unobserved heterogeneity in risk aversion is greater and more important (for profits and pricing) than unobserved heterogeneity in risk. This is consistent with the motivation for recent theoretical work, which emphasizes the importance of allowing for preference heterogeneity in analyzing insurance markets (Michael Landsberger and Isaac Meilijson, 1999; Michael Smart, 2000; David de Meza and David C. Webb, 2001; Bertrand Villeneuve, 2003; Bruno Jullien et al., forthcoming; Pierre-Andre Chiappori et al., forthcoming). Second, we find that unobserved risk has a strong positive correlation with unobserved risk aversion, and discuss possible interpretations of it. This is encouraging from a theoretical standpoint, as it retains a single crossing property, as we illustrate in the counterfactual exercise. This finding contrasts the negative correlation reported by Amy Finkelstein and Kathleen McGarry (2006) for the long-term care insurance market and by Mark Israel (2005) for automobile insurance in Illinois. We view these different results as cautioning against interpreting this correlation parameter outside of the context in which it is estimated. Even if risk preferences are stable across contexts, risk is not, and therefore neither is the correlation structure.

This study is related to two important strands of literature. The first shares our main goal of measuring risk aversion. Much of the existing evidence about risk preferences is based on introspection, laboratory experiments (Steven J. Kachelmeier and Mohamed Shehata, 1992; Vernon L. Smith and James M. Walker, 1993; Charles A. Holt and Susan K. Laury, 2002), data on bettors or television game show participants (Robert Gertner, 1993; Andrew Metrick, 1995; Bruno Jullien and Bernard Salanie, 2000; Roel M. Beetsma and Peter C. Schotman, 2001; Matilde Bombardini and Francesco Trebbi, 2005), answers given by individuals to hypothetical survey questions (W. Kip Viscusi and William N. Evans, 1990; William N. Evans and W. Kip Viscusi, 1991; Robert B. Barsky et al., 1997; Bas Donkers et al., 2001; Joop Hartog et al., 2002), and estimates that are driven by the imposed functional form relationship between static risk-taking behavior and inter-temporal substitution.⁵ We are aware of only a few attempts to recover risk preferences from decisions of regular market participants. Atanu Saha (1997) looks at firms' production decisions, and Raj Chetty (forthcoming) recovers risk preferences from labor supply. In the context of insurance, Cicchetti and Dubin (1994) look at individuals' decisions whether or not to insure against failure of inside telephone wires. Compared to their setting, in our setting events are more frequent and commonly observed, stakes are higher, the potential loss (the difference between the deductible amounts) is known, and the deductible choice we analyze is more immune to alternative preference-based explanations. Finally, in a recent working paper, Justin Sydnor (2006) uses data on deductible choices in homeowners insurance to calibrate a bound for the implied level of

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

risk aversion.⁶ An important difference between our paper and these papers is that they all rely on a representative individual framework, and therefore focus only on the level of risk aversion.⁷ In contrast, we explicitly model observed and unobserved heterogeneity in risk aversion as well as in risk. We can therefore provide results regarding the heterogeneity in risk preferences and its relationship with risk, which have potential important implications to welfare and policy. A representative individual framework cannot address such questions.

The second strand of related literature is the recent empirical literature on adverse selection in insurance markets. Much of this literature addresses the important question of whether adverse selection exists in different markets. As suggested by the influential work of Pierre-Andre Chiappori and Bernard Salanie (2000), it uses “reduced form” specifications to test whether, after controlling for observables, accident outcomes and coverage choices are significantly correlated (Georges Dionne and Charles Vanasse, 1992; Robert Puelz and Arthur Snow, 1994; John Cawley and Tomas Philipson, 1999; Amy Finkelstein and James Poterba, 2004; Finkelstein and McGarry, 2006). Alma Cohen (2005) applies this test to our data and finds evidence consistent with adverse selection. As our main goal is quite different, we take a more structural approach. By assuming a structure for the adverse selection mechanism, we can account for it when estimating the distribution of risk preferences. While the structure of adverse selection is assumed, its relative importance is 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. As we discuss in Section 3.3, this approach is conceptually similar to that of James H. Cardon and Igal Hendel (2001), who model health insurance choices and also allow for two dimensions of unobserved heterogeneity.⁸

The rest of the paper is organized as follows. Section 2 describes the environment, the setup, and the data. Section 3 lays out the theoretical model and the related econometric model, and describes its estimation and identification. Section 4 describes the results. We first provide a set of reduced form estimates, which motivate the more structural approach. We then present estimates from the benchmark specification, as well as estimates from various extensions and robustness tests. We discuss and justify some of the modeling assumptions and perform counterfactual analysis as a way to illustrate the implications of the results to profits and pricing. Section 5 concludes by discussing the relevance of the results to other settings.

⁶The possibility of using deductibles to make inferences about risk aversion was first pointed out by Jacques H. Dreze (1981). However, Dreze suggests relying on the optimality of the observed contracts (“supply side” information), while we rely on individuals’ choices of deductibles (“demand side” information).

⁷An exception is Syngjoo Choi et al. (2006), who use a laboratory experiment and, similar to us, find a high degree of heterogeneity in risk attitudes across individuals.

⁸In an ongoing project, Pierre-Andre Chiappori and Bernard Salanie (2006) estimate an equilibrium model of the French auto insurance market, where their model of the demand side of the market is conceptually similar to the one we estimate in this paper.

2 Data

2.1 Economic environment and data sources

We obtained data from a single insurance company that operates in the market for automobile insurance in Israel. The data contain information about *all* 105,800 new policyholders who purchased (annual) policies from the company during the first five years of its operation, from November 1994 to October 1999. Although many of these individuals stayed with the insurer in subsequent years, we focus through most of the paper on deductible choices made by individuals in their *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 (Alma Cohen, 2003 and 2005).

The company studied was the first company in the Israeli auto insurance 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 and appear to have a significant cost advantage (J. David Cummins and Jack L. Van Derhei, 1979). The studied company estimated that selling insurance directly results in a cost advantage of roughly 25 percent 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 because the product does not provide the “amenity” of having an agent to work with and turn to (Stephen P. D’Arcy and Neil A. Doherty, 1990). This aspect of the company clearly makes the results of the paper applicable only to those consumers who seriously consider buying direct insurance; Section 4.4.5 discusses this selection in more detail.

While we primarily focus on the demand side of the market by modeling the deductible choice, the supply side (pricing) will be relevant for any counterfactual exercise, as well as for understanding the viability of the outside option (which we do not observe and do not model). During the first two years of the company’s operations, the prices it offered were lower by about 20 percent than those offered by other, conventional insurers. Thus, due to its differentiation and cost advantage, the company had market power with respect to individuals who were more sensitive to price than to the disamenity of not having an agent. This makes monopolistic screening models apply more naturally than competitive models of insurance (e.g., Michael Rothschild and Joseph E. Stiglitz, 1976). During the company’s third year of operation (December 1996 to March 1998) it faced more competitive conditions, when the established companies, trying to fight off the new entrant, lowered the premiums 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.⁹

For each policy, our data set includes *all* the insurer’s information about the characteristics of the policyholder: demographic characteristics, vehicle characteristics, and details about his driving

⁹During 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 a strong position in the market, these two new companies did not affect pricing policies much until the end of our observation period. Right in the end of this period, the studied company acquired one of these entrants.

experience. The appendix provides a list of variables and their definitions, and Table 1 provides summary statistics. In addition, our data include the individual-specific menu of four deductible and premium combinations that the individual was offered (see below), 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.¹⁰ Finally, we use the zip codes of the policyholders’ home addresses¹¹ to augment the data with proxies for individuals’ wealth based on the Israeli 1995 census.¹²

The policies offered by the insurer (as are all policies offered in the studied market) are one-period policies with no commitment on the part of either the insurer or the policyholder.¹³ The policy resembles the U.S. version of “comprehensive” insurance. It is not mandatory, but it is held by a large fraction of Israeli car owners (above 70 percent, according to the company’s executives). The policy does not cover death or injuries to the policyholder or to third parties, which are insured through a separate mandatory policy. Insurance policies for car audio equipment, windshield, replacement car, and towing services are structured and priced separately. Certain coverages do not carry a deductible and are therefore not used in the analysis.¹⁴

Throughout the paper, we use and report monetary 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 U.S. dollars monotonically increased from 3.01 in 1995 to 4.14 in 1999 (on average, it was 3.52).¹⁵ Annual inflation was about eight percent on average, and cumulative inflation over the observation period was 48 percent. We will account for these effects, as well as other general trends, by using year dummy variables throughout the analysis.

¹⁰Throughout 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.

¹¹The 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.

¹²The Israeli Central Bureau of Statistics (CBS) associates each census respondent with a unique “statistical area,” each including between 1,000 and 10,000 residents. We matched these census tracts with zip codes based on street addresses, and constructed variables at the zip code level. These constructed variables are available for more than 80 percent of the policyholders. As a proxy for wealth, we use (gross) monthly income, which is based on self-reported income by census respondents augmented (by the CBS) with social security data.

¹³There 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 (Georges Dionne and Pierre Lasserre, 1985; Russell Cooper and Beth Hayes, 1987; Georges Dionne and Neil A. Doherty, 1994; Igal Hendel and Alessandro Lizzeri, 2003). Although such policies are observed in certain countries (Dionne and Vanasse, 1992), many insurance markets, including the one we study, use only one-period no-commitment policies (Howard Kunreuther and Mark V. Pauly, 1985).

¹⁴These include auto theft, total loss accidents, and not “at fault” accidents.

¹⁵PPP figures were about 10 percent lower than the nominal exchange rates, running from 2.60 in 1995 to 3.74 in 1999.

2.2 The menu of deductibles and premiums

Let x_i be the vector of characteristics individual i reports to the insurance company. After learning x_i , the insurer offered individual i a menu of four contract choices. One option 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 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 regular deductible level was directly linked to the regular premium according to

$$d_{it} = \min\left\{\frac{1}{2}p_{it}, cap_t\right\} \quad (1)$$

That is, it was set to one-half of the regular premium, subject to a deductible cap, cap_t , which varied over time but not across individuals. The premiums associated with the other options were computed by multiplying p_{it} by three different constants: 1.06 for “low” deductible, 0.875 for “high” deductible, and 0.8 for “very high” deductible. The regular deductible, d_{it} , was converted to the other three deductible levels in a similar way, using multipliers of 0.6 for low, 1.8 for high, and 2.6 for very high.

There are two main sources of exogenous variation in prices. The first arises from company experimentation. The multipliers described above were fixed across individuals and over time for most of the observation period, but 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 slightly modified multipliers.¹⁶ This modified formula covers almost ten percent of the sample. The second source of variation arises from discrete adjustments to the uniform cap. The cap varied over time due to inflation, competitive conditions, and as the company gained more experience (Figure 1). The cap was binding for about a third of the policyholders in our data. All these individuals would be affected by a change in the cap. Much of the variation of menus in the data 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. Indeed, this assumption holds in the data with respect to observables: there is no distinguishable difference in the distribution of observable characteristics of consumers who buy insurance just before a change in the deductible cap and those who buy just after it.

2.3 Summary statistics

The top of Table 2A summarizes the deductible menus, all are calculated according to the formula described above. Only one percent of the policyholders chose the high or very high deductible options. Therefore, for the rest of the analysis we only focus on the choice between regular and low

¹⁶For individuals with low levels of regular premiums during the specified period, the regular deductible was set at 53 percent (instead of 50 percent) of the regular premium, the low deductible was set at 33 percent (instead of 30 percent) of the regular premium, and so on.

deductible options (chosen by 81.1 and 17.8 percent of the individuals, respectively).¹⁷ Focusing only on these options does not create any selection bias because we do not omit 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, which implies that conditional on choosing high or very high deductibles, an individual would almost always prefer the regular over the low deductible.

The bottom of Table 2A, as well as Table 2B, presents summary statistics for the policy realizations. 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 4.5.1). For all these claims, the gain from choosing a low deductible is the same in the event of a claim and is equal to the difference between the two deductible levels. Therefore, the claim amount is rarely relevant for the deductible choice (and, likewise, for the company’s pricing decision we analyze in Section 4.6).

Averaging over all individuals, the annual claim rate was 0.245. One can clearly observe some initial evidence of adverse selection. On average, individuals who chose low deductible had higher claim rates (0.309) than those who chose the regular deductible (0.232). Those who chose high and very high deductibles had much lower claim rates (0.128 and 0.133, respectively). These figures can be interpreted 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 is not binding, which is the case for about two thirds of the sample, this ratio is directly given by the pricing formula and is equal to 0.3. Thus, any individual with a claim rate higher than 0.3 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 (1.3 percent according to the benchmark estimates reported below). This observation is in sharp contrast to other types of insurance contracts, such as appliance warranties, which are much more expensive than the actuarially fair price (Rabin and Thaler, 2001).

3 The empirical model

3.1 A model of deductible choice

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 annual claim rate, λ_i . Through most of the pa-

¹⁷The small frequency of “high” and “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. Considering these options, or the option of not buying insurance, creates a sharp lower bound on risk aversion for the majority of the observations, making the estimates much higher. However, given that these options are rarely selected, it is not clear to us whether they were regularly mentioned during the insurance sales process, rendering their use somewhat inappropriate.

per, we assume that λ_i is known to the individual. We also assume that λ_i is independent of the deductible choice, i.e., that there is no moral hazard. Finally, we assume that, in the event of an accident, the value of the claim is greater than d_i^h . We revisit all these assumptions in Sections 4.4 and 4.5. 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. Both the premium and the probability of a claim are proportional to the length of the time interval taken into account, so it is convenient to think of the contract choice as a commitment for only a short amount of time. This approach has several advantages. First, it helps to account for early cancellations and truncated policies, which together constitute thirty percent of the policies in the data.¹⁸ Second, it makes the deductible choice independent of other longer-term uncertainties faced by the individual, so we can focus on static risk-taking behavior. Third, this formulation helps to obtain a simple framework for analysis, which is attractive both analytically and computationally.¹⁹

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

$$v(p, d) \equiv (1 - \lambda t)u(w - pt) + (\lambda t)u(w - pt - d) \quad (2)$$

We characterize the set of parameters that will make the individual indifferent between the two offered contracts. This set provides a lower (upper) bound on the level of risk aversion for individuals who choose the low (high) deductible (for a given λ). Thus, we analyze the equation $v(p^h, d^h) = v(p^l, d^l)$. By taking limits with respect to t (and applying L'Hopital's rule), we obtain

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

or

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

¹⁸As can be seen in Table 2A, 70 percent of the policies are observed through their full duration (one year). About 15 percent are truncated by the end of our observation period, and the remaining 15 percent 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.

¹⁹This 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 a 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 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.

In our benchmark specification, we assume that the third derivative of the vNM utility function is not too large. A Taylor expansion for both terms on the right-hand side of equation (4), i.e., $u(w - d) \approx u(w) - du'(w) + \frac{d^2}{2}u''(w)$, implies that

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

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

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

or

$$r \equiv \frac{-u''(w)}{u'(w)} \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}} \quad (7)$$

where r is the coefficient of absolute risk aversion at wealth level w .

Equation (7) defines an indifference set in the space of risk and risk aversion, which we will refer to by $(r^*(\lambda), \lambda)$ and $(\lambda^*(r), r)$ interchangeably. Both $r^*(\lambda)$ and $\lambda^*(r)$ have a closed-form representation, a property that will be computationally attractive for estimation.²⁰ Both terms are individual specific, as they depend on the deductible menu, which varies across individuals. For the rest of the paper, we regard each individual as associated with a two-dimensional type (r_i, λ_i) . An individual with a risk parameter λ_i , who is offered a menu $\{(p_i^h, d_i^h), (p_i^l, d_i^l)\}$, will choose the low-deductible contract if and only if his coefficient of absolute risk aversion satisfies $r_i > r_i^*(\lambda_i)$. Figure 2 presents a graphical illustration.

3.2 The benchmark econometric model

3.2.1 The econometric model

The econometric model we estimate is fully described by the five equations presented in this section. Our objective is to estimate the joint distribution of (λ_i, r_i) – the claim rate and coefficient of absolute risk aversion – in the population of policyholders, conditional on observables x_i . The benchmark formulation assumes that (λ_i, r_i) follows a bivariate lognormal distribution. Thus, we can write the model as

$$\ln \lambda_i = x_i' \beta + \varepsilon_i \quad (8)$$

$$\ln r_i = x_i' \gamma + v_i \quad (9)$$

with

$$\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_r \\ \rho \sigma_\lambda \sigma_r & \sigma_r^2 \end{bmatrix} \right) \quad (10)$$

²⁰For example, estimating the CARA version of the model (Section 4.4), for which $r^*(\lambda)$ does not have a closed-form representation, takes almost ten-times longer.

Neither λ_i nor r_i is directly observed. Therefore, we treat both as latent variables. Loosely speaking, they can be thought of as random effects. We observe two variables, the number of claims and the deductible choice, which are related to these two unobserved components. Thus, to complete our econometric model we have to specify the relationship between the observed variables and the latent ones. This is done by making two structural assumptions. First, we assume that the number of claims is drawn from a Poisson distribution, namely

$$claims_i \sim Poisson(\lambda_i t_i) \quad (11)$$

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

$$\Pr(choice_i = 1) = \Pr\left(r_i > \frac{\frac{\Delta p_i}{\lambda_i \Delta d_i} - 1}{\bar{d}_i}\right) = \Pr\left(\exp(x_i' \gamma + v_i) > \frac{\frac{\Delta p_i}{\exp(x_i' \beta + \varepsilon_i) \Delta d_i} - 1}{\bar{d}_i}\right) \quad (12)$$

With no unobserved heterogeneity in risk ($\varepsilon_i = 0$), equation (12) reduces to a simple probit. In such a case, one can perfectly predict risk from the data, denote it by $\hat{\lambda}(x_i)$, and construct an additional covariate, $\ln\left(\frac{\Delta p_i / (\hat{\lambda}(x_i) \Delta d_i) - 1}{\bar{d}_i}\right)$. Given the assumption that risk aversion is distributed log-normally, running the above probit regression and renormalizing the coefficient on the constructed covariate to -1 (instead of the typical normalization of the variance of the error term to 1) has a structural interpretation, with $\ln r_i$ as the dependent variable. However, Cohen (2005) provides evidence of adverse selection in the data, implying the existence of unobserved heterogeneity in risk. This makes the simple probit regression misspecified. Estimation of the full model is more complicated. Once we allow for unobserved heterogeneity in both unobserved risk aversion (v_i) and claim rate (ε_i), we have to integrate over the two-dimensional region depicted in Figure 2 for estimation.

3.2.2 Estimation

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

$$L(claims_i, choice_i | \theta) = \Pr(claims_i, choice_i | \lambda_i, r_i) \Pr(\lambda_i, r_i | \theta) \quad (13)$$

where θ is a vector of parameters to be estimated. While formulating the empirical model using likelihood may help our 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 complex integral for each individual in the data. In contrast, Markov Chain Monte Carlo (MCMC) Gibbs sampling is quite attractive in such a case. Using data augmentation of latent variables (Martin A. Tanner and Wing Hung 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 sampling from truncated normal distributions, which is significantly less

computationally demanding (e.g., Luc Devroye, 1986). This feature, combined with the idea of a “sliced sampler” (Paul 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 lognormality assumption implies that $F(\ln \lambda_i | r_i)$ and $F(\ln r_i | \lambda_i)$ follow a (conditional) normal distribution, allowing us to restrict attention to univariate draws, further reducing the computational burden.

The appendix provides a full description of the 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 less standard 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. The final step is to sample from the posterior distribution of $\ln \lambda_i$, 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 number of claims, which provides additional information about the posterior of λ_i . Thus, the posterior for λ_i takes an unfamiliar form, for which we use a “sliced sampler.”

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. The results are based on the posterior mean and posterior standard deviation from these 90,000 draws. Note that each iteration involves generating separate draws of (λ_i, r_i) for each individual. Performing 100,000 iterations of the benchmark specification (coded in Matlab) takes about 60 hours on a Dell Precision 530 workstation.

3.3 Identification

The parametric version of the model is identified mechanically. There are more equations than unknowns and no linear dependencies among them, so (as also verified using Monte Carlo exercises) the model parameters can be backed out from simulated data. Our goal in this section is *not* to provide a formal identification proof. Rather, we want to provide intuition for which features of the data allow us to identify particular parameters of the model. The discussion also highlights the assumptions that are essential for identification vis-a-vis those that are only made for computational convenience (making them, in principle, testable).

Discussion of non-parametric identification The main difficulty in identifying the model arises from the gap between the (ex ante) risk type, λ_i , which individuals use when choosing a deductible, and the (ex post) realization of the number of claims we observe. We identify between the variation in risk types and the variation in risk realizations using our parametric distributional assumptions. The key is that the distribution of risk types can be uniquely backed out from the claim data alone. This allows us to use the deductible choice as an additional layer of information, which identifies unobserved heterogeneity in risk aversion.²¹ Any distributional assumption that

²¹Cardon and Hendel (2001) face a similar identification problem in the context of health insurance. They use variation in coverage choice (analogous to deductible choice) to identify the variation in health-status signals (analo-

allows us to uniquely back out the distribution of risk types from claim data would be sufficient to identify the distribution of risk aversion. As is customary in the analysis of count processes, we make a parametric assumption, that claims are generated by a lognormal mixture of Poisson distributions (Section 4.4 discusses this further and explores an alternative). Using a mixture enables us to account for adverse selection through unobserved heterogeneity in risk. It also allows us to better fit the tails of the claim distribution. In principle, a more flexible mixture or a more flexible claim-generating process could be identified, as long as the claims data is sufficiently rich.²²

Once the distribution of risk types is identified from claims data, the marginal distribution of risk aversion (and its relationship to the distribution of risk types) is non-parametrically identified from the variation in the offered menus discussed in Section 2. This variation implies different deductible and premium options to identical (on observables) individuals who purchased insurance at different times. Different menus lead to different indifference sets (similar to the one depicted in Figure 2). These sets often cross each other and non-parametrically identify the distribution of risk aversion and the correlation structure, at least within the region in which the indifference sets vary. For the tails of the distribution, as is typically the case, we have to rely on parametric assumptions or use bounds. The parametric assumption of lognormality we use for most of the paper is only made for computational convenience.

Intuition for the parametric identification mechanism Variation in the offered menus is important for the non-parametric identification. The parametric assumptions could identify the model without such variation. Thus, to keep the intuition simple, let us take the bivariate lognormal distribution as given and, contrary to the data, assume that all individuals are identical on observables and face the same menu. Suppose also that all individuals are observed for exactly one year and have up to two claims.²³ In this simplified case, the model has five parameters to be estimated – the mean and variance of risk, μ_λ and σ_λ^2 , the mean and variance of risk aversion, μ_r and σ_r^2 , and the correlation parameter, ρ – and 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 a low deductible within each “claim group.” Given our distributional assumption about the claim-generating process, we can use α_0 and α_1 to uniquely identify μ_λ and σ_λ^2 . Loosely, μ_λ is identified by the average claim rate in the data and σ_λ^2 is identified by how fat the tail of the claim distribution is, i.e., by how large $\frac{\alpha_2}{\alpha_1}$ is compared to $\frac{\alpha_1}{\alpha_0}$. Given μ_λ and σ_λ^2 and the lognormality assumption, we can (implicitly) construct

gous to risk types) from the variation in health expenditure (analogous to number of claims). They can rely on the coverage choice to identify this because they make an assumption regarding unobserved heterogeneity in preferences (i.i.d. logit). We take a different approach, as our main goal is to estimate, rather than assume, the distribution of preferences.

²²Although it may seem that the claim data is limited (as they only take integer values between 0 and 5 in our data), variation in policy durations generates continuous variation in the observed claim propensity. Of course, this variation also introduces an additional selection into the model due to policy cancellations, which are potentially endogenous. The results are similar when we only use expired and truncated policies.

²³This variation is sufficient (and necessary) to identify the benchmark model. The data provide more variation: we observe up to five claims per individual, we observe continuous variation in the policy durations, we observe variation in prices, and we exploit distributional restrictions across individuals with different observable characteristics.

a posterior distribution of risk types for each claim group, $F(\lambda|r, \text{claims} = c)$, and integrate over it when predicting the deductible choice. This provides us with three additional moments, each of the form

$$E(\varphi_c) = \int \int \Pr(\text{choice} = 1|r, \lambda) dF(\lambda|r, \text{claims} = c) dF(r) \quad (14)$$

for $c = 0, 1, 2$. These moments identify the three remaining parameters of the model, μ_r , σ_r^2 , and ρ .

Let us now provide more economic content to the identification argument. Using the same example, and conditional on identifying μ_λ and σ_λ^2 from the claim data, one can think about the deductible choice data, $\{\varphi_0, \varphi_1, \varphi_2\}$, as a graph $\varphi(c)$. The absolute level of the graph identifies μ_r . In the absence of correlation between risk and risk aversion, the slope of the graph identifies σ_r^2 ; with no correlation, the slope should always be positive (due to adverse selection), but higher σ_r^2 would imply a flatter graph because more variation in the deductible choices will be attributed to variation in risk aversion. Finally, ρ is identified by the curvature of the graph. The more convex (concave) the graph is, the more positive (negative) is the estimated ρ . For example, if $\varphi_0 = 0.5$, $\varphi_1 = 0.51$, and $\varphi_2 = 0.99$, it is likely that σ_r^2 is high (explaining why φ_0 and φ_1 are so close) and ρ 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 the identification of ρ relies on observing individuals with multiple claims (or different policy durations) and that it is likely to be sensitive to the distributional assumptions. The data (Table 2B) provide a direct (positive) correlation between deductible choice and claims. The structural assumptions allow us to explain how much of this correlation can be attributed to adverse selection. The remaining correlation (positive or negative) is therefore attributed to correlation in the underlying distribution of risk and risk aversion.

4 Results

4.1 Reduced form estimation

To get an initial sense for the levels of absolute risk aversion implied by the data, we use a simple back-of-the-envelope exercise. We compute unconditional averages of Δp , Δd , λ , d^h , d^l , and \bar{d} (Table 2A),²⁴ and substitute these values in equation (7). The implied coefficient of absolute risk aversion is $2.9 \cdot 10^{-4} \text{ NIS}^{-1}$.²⁵ This figure could be thought of as the average indifference point, implying that about 18 percent of the policyholders have coefficients of absolute risk aversion exceeding it. To convert to U.S. dollar amounts, one needs to multiply these figures by the average exchange rate (3.52), resulting in an average indifference point of $1.02 \cdot 10^{-3} \text{ } \US^{-1} . This figure is less than half of a similar estimate reported by Sydnor (2006) for buyers of homeowners insurance, but about three and thirteen times higher than comparable figures reported by Gertner (1993) and Metrick (1995), respectively, for television game show participants.

²⁴The unconditional λ is computed by maximum likelihood, using the data on claims and observed durations of the policies.

²⁵Using the CARA specification, as in equation (16), we obtain a slightly lower value of $2.5 \cdot 10^{-4} \text{ NIS}^{-1}$.

Table 3 provides reduced form analysis of the relationship between the observables and our two left-hand side variables, the number of claims and the deductible choice.²⁶ Column (1) reports the estimates from a Poisson regression of the number of claims on observed characteristics. This regression is closely related to the risk equation we estimate in the 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 a strong predictor of future claims. Young drivers are 50-70 percent 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.²⁷ This decline might also be partly due to the better selection of individuals the company obtained over time as it gained more experience (Cohen, 2003).

Column (2) and (3) of Table 3 present estimates from probit regressions in which the dependent variable is equal to 1 if the policyholder chose a low deductible, and is equal to 0 otherwise. Column (3) shows the marginal effects of the covariates on the propensity to choose a low deductible. These marginal effects do not have a structural interpretation, as the choice of low deductible depends on its price, on risk aversion, and on risk. In this regression we again observe a strong trend over time. Fewer policyholders 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 option.²⁸ The effect of other covariates will be discussed later in the context of the full model. In unreported probit regressions, we also test the qualitative assumptions of the structural model by adding three additional regressors, the price ratio $\frac{\Delta p_i}{\Delta d_i}$, the average deductible offered \bar{d}_i , and the risk rate, $\hat{\lambda}(x_i)$, as predicted from the Poisson regression of column (1). All three additional regressors enter with the predicted sign, and with large and highly significant marginal effects.²⁹

Finally, column (2) of Table 3 presents an important specification of the probit regression, in which $\ln\left(\frac{\Delta p_i / (\hat{\lambda}(x_i) \Delta d_i) - 1}{\bar{d}_i}\right)$ is added as a regressor, and its coefficient is normalized to -1 .³⁰ As

²⁶We find positive correlation in the errors of these two regressions, suggesting the existence of adverse selection in the data and motivating a model with unobserved heterogeneity in risk. This test is similar to the bivariate probit test proposed by Chiappori and Salanie (2000) and replicates earlier results reported in Cohen (2005).

²⁷In particular, traffic fatalities and counts of traffic accidents in Israel fell by 11 percent and 18 percent during 1998 and 1999, respectively.

²⁸Such 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 were related to observables (e.g., education), the coefficients on such observables would be biased upwards. This is not a major concern, given that the coefficients on most of the covariates are fairly stable when we estimate the benchmark model separately for different years.

²⁹The estimated marginal effect (z -statistic in parentheses) is -0.352 (-13.76), $1.6 \cdot 10^{-4}$ (14.81), and -0.154 (-2.55) for $\frac{\Delta p_i}{\Delta d_i}$, \bar{d}_i , and $\hat{\lambda}(x_i)$, respectively.

³⁰The level is a normalization. The sign is estimated. Had the sign on this regressor been positive, this would have implied a rejection of the model.

already mentioned, if we assume that there is no unobserved heterogeneity in risk, then column (2) is analogous to the $\ln r_i$ equation of the benchmark model. This restriction of the model is rejected by the data. About ten percent of the individuals are predicted to have $\hat{\lambda}(x_i) > \frac{\Delta p_i}{\Delta d_i}$, implying a choice of low deductible for *any* level of risk aversion. However, many of these individuals still choose higher deductible levels. Column (2) reports the results for the remaining individuals (94,000 out of 105,800) for which the regression can be run. While the signs of the estimated coefficients are similar to those in the benchmark model presented below, the restricted version of the model suggests much higher levels and dispersion of risk aversion, well above any reasonable level.³¹ The full estimation of the benchmark model clearly rejects this restriction on the model.

4.2 Estimation of the benchmark model

4.2.1 The effect of individual characteristics on risk aversion

Table 4 presents the estimation results from the benchmark model. The second column shows how the level of absolute risk aversion is related to individual characteristics. As the dependent variable is in natural logarithm, coefficients on dummy variables can be directly interpreted as approximate percentage changes. The vast majority of these coefficients are quite stable across a wide range of specifications, which are mentioned later.

The results indicate that women are more risk averse than men, and have a coefficient of absolute risk aversion about 20 percent greater than that of men. These results are consistent with those of Donkers et al. (2001) and Hartog et al. (2002). The estimated effect of age suggests a non-monotone pattern of risk preferences over the life cycle. The estimated coefficients imply that initially (that is, at age 18, the youngest individual in the data) individuals become less risk averse with age, but around the age of 48, individuals start becoming more risk averse.³² Married individuals are estimated to be significantly more risk averse compared to singles, while divorced individuals are less (although the coefficient is statistically insignificant).

The analysis suggests that variables that are likely to be correlated with income or wealth, such as post high school education and the value of the car, have a positive coefficient, indicating that wealthier people have higher levels of absolute risk aversion. Although we do not have data on individuals' income or wealth, we use other proxies for income in additional specifications and obtain mixed results. When we include as a covariate the average household income among those who live in the same zip code, we obtain a significant and negative coefficient of -0.333 (0.154) (standard deviation in parentheses). When we match zip code income on demographics of the individuals, the coefficient is effectively zero, -0.036 (0.047), and when we use a socioeconomic

³¹For the implied median level of risk aversion, the restricted model produces a similar estimate to the estimate we report below for the benchmark model. However, since some individuals who chose the low deductible are estimated to have very low claim rates, the restricted model is “forced” to estimate very high risk aversion for these individuals (in contrast, the benchmark model can explain such choices by positive risk residuals), resulting in very high dispersion and (due to the lognormality assumption) very high average risk aversion, which is about 10^{25} higher than the benchmark estimates we report below.

³²This non-monotone pattern may explain why age enters with different signs in the estimation results of Donkers et al. (2001) and Hartog et al. (2002). A somewhat similar U-shape pattern with respect to age is also reported by Sumit Agarwal et al. (2006) in the context of consumer credit markets.

index of the locality in which the individual lives, the coefficient is positive, 0.127 (0.053).³³ Thus, overall we interpret our findings as suggestive of a non-negative, and perhaps positive, association between income/wealth and absolute risk aversion.

At first glance, these results may appear to be inconsistent with the widely held belief that absolute risk aversion declines with wealth. One should distinguish, however, 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. Testing this property would require observing the same individual making multiple choices at different wealth levels. Rather, our results indicate that individuals with greater wealth have utility functions that involve a greater degree of risk aversion. It might be that wealth is endogenous and that risk aversion (or unobserved individual characteristics that are correlated with it) leads individuals to save more, to obtain more education, or to take other actions that lead to greater wealth.³⁴

Let us make several additional observations. First, while 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 does not constrain the relationship between the coefficients in the two equations. Rather, it is the data that “speak up.” Second, 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. Third, policyholders who tend to use the car for business are less risk averse. This could be because uninsured costs of accidents occurring to such policyholders are tax deductible. Fourth, policyholders who reported three full years of claims 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 three full years of claims 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, we find a strong trend towards lower levels of risk aversion over time. This is a replication of the probit results discussed earlier.

³³The full results from these specifications are provided in the online appendix. The reason we do not use these income variables in the benchmark specification is their imperfect coverage, which would require us to omit almost twenty percent of the individuals. Other results from these regressions are similar to those we report for the benchmark specification.

³⁴One may be tempted to interpret the positive wealth effects as an indirect indication of 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, the insurance company observes these proxies for wealth and conditions on them when setting prices. Since the willingness to pay for insurance is likely to be correlated with the willingness to pay for the additional insurance provided by the low deductible option, premiums already reflect this variation. We condition on the level of the premium. Second, paying less ex ante implies paying more ex post, so applying the credit constraint argument only to the ex ante payment but not to the probabilistic ex post deductible payments has no theoretical foundation. Essentially, the setup of the model links the ex ante decisions to the ex post losses, which are both driven by the curvature of the vNM utility function.

4.2.2 The effect of individual characteristics on claim risk

The first column of Table 4 describes the relationship between observables and risk exposure. Accident risk is higher for divorced individuals and lower for people with academic education. Bigger, older, more expensive, and non-commercial cars are all more likely to be involved in an accident. Driving experience reduces accident rates, as do measures of less intense use of the car, while young drivers are more exposed to risk. Claim propensity is highly correlated over time: the voluntary report of past claims is a strong predictor of future claims. This risk equation 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. The similarity between these two sets of results is to be expected, as the risk regression is mainly identified from the data on claims, so incorporating the information on deductible choice does not qualitatively change the conceptual identification strategy (see Section 3.3). If the results were not similar, this could have indicated a misspecification of the model. The slight differences between the risk regressions in Table 3 and Table 4 are mainly driven by the structural assumptions. First, the benchmark model estimates a lognormal mixture of Poisson rates, 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.

4.2.3 The implied level of risk aversion

One of the main goals of the estimation is to obtain measures of the level of risk aversion in the population we study. Since we use Gibbs sampling and augment the latent coefficients of absolute risk aversion, we can directly obtain the posterior distribution of various moments of this distribution. At each iteration of the Gibbs sampler, we compute the mean, standard deviation (across individuals), and various percentiles of the simulated draws of λ_i and r_i and the correlation between them. The right-most column of Table 4 reports the averages and standard deviations of these computed quantities over the iterations of the Gibbs sampler.³⁵ The implied risk aversion of the mean individual is 0.0019, which is about seven times greater than the back-of-the-envelope calculation presented in the previous section. As we assume a lognormal distribution and estimate a relatively high dispersion coefficient, σ_r , the estimates also imply significantly lower values for the median level of risk aversion. As shown later, the qualitative pattern of these results is quite robust across specifications.

Table 5 summarizes all our findings regarding the level and dispersion of risk aversion. It presents two ways to interpret the estimates we obtained from the benchmark model, as well as comparisons to specifications with CARA utility and with incomplete-information (discussed later), and to other comparable figures in the literature. Our benchmark estimate suggests that an average

³⁵Note that these estimated quantities are unconditional. In computing these quantities, we integrate over the distribution of observable characteristics in the data, so one cannot compute these estimates from the estimated parameter directly.

quadratic utility maximizer³⁶ will be indifferent about participating in a fifty-fifty lottery in which he gains 100 dollars or loses 56.05 dollars. A CARA specification suggests a significantly lower value for average risk aversion, making the mean individual indifferent about a fifty-fifty lottery of gaining 100 dollars or losing 76.51 dollars. The results from the incomplete information model for the mean level of risk aversion are in between these estimates. All the results suggest that, although the mean individual exhibits a significant level of risk aversion, heterogeneity in risk preferences is important, and the median individual is almost risk neutral with respect to lotteries of 100 dollar magnitude. Although the mean is always greater than the median under the lognormal distribution, the large 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).

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, both of these papers measure risk aversion for television game show participants; these are highly selected groups in a rather “risk friendly” environment.³⁷ Second, the magnitudes of the stakes are higher. The show participants bet on several thousand dollars and more, while our average individual risks much lower stakes, in 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. We discuss this further in Section 5.

A different way to quantify our estimate is by reporting them in relative terms. There is no consensus in the literature as to the relevant wealth that is taken into account in such decisions. Therefore, for comparability, we follow the closest papers in this respect (e.g., Gertner, 1993), and use annual income as the relevant wealth. We multiply the estimated coefficient of absolute risk aversion by the average annual income in Israel during the observation period. Under the (questionable) 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 Table 5 indicates, our benchmark specification results in an implied coefficient of relative risk aversion of about 97. A CARA specification results in a lower coefficient of 44. On the other hand, the median estimate for relative risk aversion is well below one. Thus, the widely used estimate of a low single-digit coefficient of relative risk aversion falls somewhere between the median and the mean, and between the median and the 75th percentile of the risk aversion distribution.

4.2.4 The relationship between unobserved risk and unobserved risk aversion

Table 4 allows us to make observations about the relationship between risk and risk aversion. We first discuss the relative importance of unobserved heterogeneity of both dimensions. In the population we study, unobserved heterogeneity in risk aversion (σ_r) is much greater than unobserved heterogeneity in risk (σ_λ). This is true both in absolute terms (3.15 compared to 0.15, respectively)

³⁶For such an individual, the second-order Taylor expansion we use in Section 3.1 is exact.

³⁷We 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.

and after normalizing by the corresponding mean level,³⁸ using the coefficient of variation as a measure of dispersion (0.27 compared to 0.1, respectively). It is also true for the overall, unconditional dispersion. This could indicate that selection on risk aversion is more important in our data than adverse selection.³⁹ However, 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. Profits, for example, are affected directly by risk but not by risk aversion, so the above comparison could be misleading. Therefore, we relegate the discussion of this issue to Section 4.6, where we show that even when we look at pricing and profits, heterogeneity in risk aversion is more important.

Table 4 also indicates a strong and significant positive correlation of 0.84 between unobserved risk aversion and unobserved risk. This result may be 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 (2006) supports such intuition by documenting a negative correlation between risk aversion and risk in the market for long-term care insurance (see also Israel (2005) for the auto insurance market in Illinois). Our market, however, might be special in ways that could produce a positive correlation. First, in contrast to most insurance markets, where a policyholder’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. 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. The intensity of vehicle use, for example, might be an 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,⁴⁰ controlling for

³⁸All covariates are measured in deviations from their sample mean, so the estimated constant in each equation is the estimated mean of the left-hand side variable.

³⁹An additional observation is that, given our estimates, observables account for slightly less than 50 percent of the variation in $\ln r_i$, but for almost 65 percent of the variation in $\ln \lambda_i$. This may seem surprising given the finding that dispersion in risk aversion is more important and thus should be the focus of insurance companies. However, this finding is consistent with the conventional wisdom that insurance companies spend much effort and resources on collecting information that helps in risk classification, but only little effort on information that predicts willingness to pay.

⁴⁰Insurance companies typically do not use these 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. Indeed the estimated elasticity of risk with respect to estimated mileage is 0.06 (0.012), which seems low, suggesting a bias downwards, potentially due to “errors in variables” bias.

annual mileage reported by policyholders reduces the estimated correlation coefficient to 0.68.⁴¹ 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. A third explanation for the positive estimated correlation is the distributional assumption. As discussed in Section 3.3, the correlation coefficient is probably the parameter that is most sensitive to these assumptions. Indeed, as discussed later, when we change our assumptions about the Poisson process and use an alternative distribution with extremely thin tails, the estimated correlation coefficient reverses signs. We, of course, view the assumptions of the benchmark model as more appropriate, and therefore maintain the view that the data suggest a positive correlation. Finally, note that while the correlation parameter we estimate is high, the implied *unconditional* correlation between risk and risk aversion is less than 0.25 across all reported specifications. This is because the coefficients on the same covariate (for example, the size of the car or whether the car is used for business) often affect risk and risk aversion in opposite directions, and because of the logarithmic transformation.

4.3 Stability of the risk aversion coefficients

Estimating risk preferences is motivated by the idea that the same (or similar) risk aversion parameter may explain risky decisions across multiple contexts. This idea is at the heart of the current debate we mention in the introduction, regarding the empirical relevance of expected utility and whether the standard construct of a single risk aversion parameter that applies across many types of decisions is the right way to think about consumer behavior. To the best of our knowledge, there is no direct empirical evidence on this issue. While we defer much of this discussion to the concluding section, we verify below that, at least over a limited set of choices in which we observe the same individuals, the estimated risk preferences help to explain choices over time and across different contexts.

Stability across contexts Ideally, one would like to show that the estimated risk preferences are stable across multiple lines of insurance, and perhaps even across other risky decisions individuals face. Absent data on such choices, we provide evidence that the estimated risk aversion coefficients help in predicting other related risky choices. Individuals in our data had to make three additional coverage choices in addition to the main policy choice that we have analyzed so far. Individuals had to choose whether to purchase coverage for car audio equipment (bought by 49.8 percent of the sample of new policyholders), for towing services (bought by 92.9 percent), and for windshield damage (bought by 95.4 percent). These coverages are sold by the same company but are administered by a third party, so we do not have information about their realization. Therefore, we cannot perform a similar exercise to the one we perform for the deductible choice. Of course, one should not expect these additional coverage choices to be perfectly correlated with the deductible choice. Even if the same risk preferences are an important factor for all of these decisions, variation in risk across coverages is also important. For example, ownership of an expensive car audio system is

⁴¹The full results from this specification are provided in the online appendix. Other results from this specification are similar to those we report for the benchmark specification.

likely to affect the purchase of audio coverage, independent of risk preferences. Similarly, owners of old cars may value more coverage for towing services. Risk preferences, however, should enter into all of these coverage choices, and therefore we expect these additional coverages to be positively correlated with the choice of a low deductible.

We coded these three additional coverage choices as three dummy variables and use them as additional covariates to verify that they can help explain unobserved risk aversion. First, we add these variables to the probit regression reported in column (3) of Table 3. They all have a positive, statistically and economically significant power in predicting the deductible choice.⁴² To verify that these reported correlations are not driven by correlation in risks, we also estimate the benchmark model with these additional variables as covariates. The estimated coefficients on the three additional coverages are positive, and two of the three are significant in both the risk and risk aversion equations: they are 0.053 (0.019), 0.0002 (0.021), and 0.057 (0.023) in the risk equation, and 0.354 (0.064), 0.123 (0.122), and 0.813 (0.174) in the risk aversion equation, for audio, towing, and windshield coverage, respectively (standard deviations in parentheses).⁴³ This suggests that the estimated risk preferences help in explaining multiple coverage choices across these related contexts.

Stability over time We now provide evidence that the estimated risk preferences are also stable over time. All the individuals who decide to renew their policy with the company after it expires are, at least in principle, free to change their deductible choice. However, an overwhelming majority of individuals (more than 97 percent) do not change their deductible choices when they renew. Even individuals who initially chose the rarely chosen “high” and “very high” deductible levels typically do not change their choices. Therefore, it is also not surprising that estimating the model on the first deductible choice of individuals and the second deductible choice of the same individuals yields similar results, as shown at the end of the next section. Of course, changes in deductible choices do not necessarily imply changes in risk preferences. Risk may also change over time, and may drive such changes. At the same time, while persistence in deductible choices over time is consistent with stable risk preferences, it may also be driven by many other factors. For example, it is reasonable to think that individuals do not devote the same amount of thought to their renewal decision as compared to their initial choice, leading to the observed persistence. This caveat is the main reason why we focus most of the analysis only on new policies.

4.4 Robustness

This section discusses the robustness of the main results. Results from all specifications we mention in this section are summarized in Table 6. The full results from these specifications are relegated to an online appendix.

⁴²They are estimated to increase, on average, the probability of choosing a low deductible by 0.019 (8.19), 0.023 (5.27), and 0.041 (7.58) for audio, towing, and windshield coverage, respectively ($z - stat$ in parentheses).

⁴³The full results from this specification are provided in the online appendix. Other results from this specification are similar to those we report for the benchmark specification.

4.4.1 The von Neumann Morgenstern utility function

To derive the benchmark model, we assume a negligible third derivative of the vNM utility function. This is attractive, as it allows us to summarize risk preferences by a one-dimensional parameter, $r \equiv \frac{u''(w)}{u'(w)}$. However, there is a large literature (e.g., Miles S. Kimball and N. Gregory Mankiw, 1989) emphasizing the importance of a (positive) third derivative of the utility function, which leads to precautionary saving. It is therefore important to address the sensitivity of the results to this assumption.

By allowing a third-order Taylor approximation of equation (4), we obtain

$$r \equiv \frac{-u''(w)}{u'(w)} \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}} - \frac{u'''(w)}{u'(w)} \frac{(d_h^2 + d_h d_l + d_l^2)}{6\bar{d}} \quad (15)$$

which reduces to equation (7) when $u'''(w) = 0$. This illustrates that a positive third derivative provides an additional (precautionary) incentive to insure. It also shows that in order to fully describe preferences in the presence of a third derivative, a one-dimensional parameter does not suffice: one needs to know both $r \equiv \frac{u''(w)}{u'(w)}$ and $\frac{u'''(w)}{u'(w)}$. This makes it less attractive for estimation.

We can allow for a positive third derivative without expanding the dimensionality of preferences by imposing a parametric functional form on the utility function. Essentially, such parametric form imposes a relationship between $\frac{u''(w)}{u'(w)}$ and $\frac{u'''(w)}{u'(w)}$. Two standard forms are those that exhibit constant absolute risk aversion (CARA) and constant relative risk aversion (CRRA). CRRA requires us to make additional assumptions about the relevant (and unobserved) wealth level of each individual, making it less attractive. The CARA case is implemented below.

With CARA utility, we substitute $u(w) = -\exp(-rw)$ in equation (4) and rearrange to obtain

$$\lambda = \frac{r \Delta p}{\exp(rd^h) - \exp(rd^l)} \quad (16)$$

This equation defines the indifference set. Unlike the benchmark model, there is no closed-form representation for $r_i^*(\lambda)$, a property that makes estimation significantly slower. Due to the precautionary saving incentive arising from the third derivative, we can use a lower level of absolute risk aversion to rationalize the low deductible choice, given λ . In other words, the CARA indifference set is flatter in comparison with the benchmark case depicted in Figure 2. Thus, the CARA specification will generally lead to lower estimates of the coefficient of absolute risk aversion, as shown in Table 6. The mean and dispersion of estimated absolute risk aversion (but not the median) are smaller by about a factor of 2. The general qualitative results, however, remain the same. As shown in Table 5, the mean individual is quite risk averse with respect to lotteries of 100 dollar magnitude, the median is close to risk neutral with respect to such lotteries, and heterogeneity in risk aversion is important.

4.4.2 The claim-generating process

In the benchmark model, we assume that claims are generated by a Poisson model. This assumption is both analytically attractive (it is important in deriving equation (4)) and, in our view, a reasonable approximation of the data generating-process for claims. The Poisson distribution has

been widely used in economics to model accident counts (e.g., Nancy L. Rose, 1990; Dionne and Vanasse, 1992; Ron Michener and Carla Tighe, 1992). In using this distribution, researchers have followed the long history of the use of Poisson by actuaries and insurance specialists going back to Ladislaus Bortkiewicz (1898).

An important restriction of the Poisson distribution is that its mean and variance are the same. Although some economic studies confirmed this assumption for particular accident data (e.g., Christopher M. Auld et al., 2001), it is often the case that this restriction is falsified. The most common deviation from the Poisson restriction is that of fat tails, i.e. variance that is higher than the mean. This led researchers to use a negative binomial distribution to accommodate this regularity by introducing a second parameter, which de-links the relationship between the mean and the variance. One natural interpretation of the fat tails is that of unobserved heterogeneity, and the negative binomial distribution can be viewed as a Gamma mixture of Poisson processes. Consistent with this view, we assume that the claim-generating process follows a Poisson process at the individual level, but allow unobserved heterogeneity in risk, and estimate a lognormal mixture of Poisson processes, which is similar to a negative binomial. The dispersion parameter we estimate, σ_λ , is a free parameter, which is identified by the fatness of the tails of the claim distribution.

While one would like to allow fat tails of the aggregate claim distribution, one may criticize the assumption of a Poisson distribution at the individual level for having tails that may be “too fat.” Recently, Jaap H. Abbring et al. (2003) and Mark Israel (2004) provided evidence for negative state dependence in data from auto insurance claims, similar to the ones we use. Controlling for heterogeneity across individuals, these papers show that a second claim is less likely than a first. This may happen due to experience rating or, perhaps, to more careful driving and less intense use of a car after an accident. The Poisson distribution assumes that the second accident is just as likely, so negative state dependence may suggest thinner tails at the individual level.

To verify that the main results are not sensitive to this restriction of the Poisson model, we perform two tests. First, we estimate the benchmark model on a sample that only includes individuals with one or no claims. The model is still identified using variation in policy durations. The estimates of the level of risk aversion and its dispersion across individuals remain similar to those of the benchmark model (Table 6). The estimates of risk, and the correlation with risk aversion, are significantly lower. But this is expected: we selected out of the sample the high-risk individuals.

As a further test for the sensitivity of the results to possible negative state dependence, we estimated the model with a different distribution at the individual level, which gives rise to much thinner tails. We could not find any known distribution of count variables that gives rise to tails thinner than Poisson’s, so we made one up. In particular, for a given Poisson rate, let p_n be the probability of observing n occurrences. The distribution we take to the data is such that

$$\Pr(n) = \frac{p_n^2}{\sum_{m=0}^{\infty} p_m^2} \tag{17}$$

Thus, this distribution makes the probability of multiple claims much lower, probably much more so than any realistic negative state dependence in the data. Essentially, such a distribution makes the econometrician view individuals with multiple claims as high-risk individuals with much more certainty, as it requires much more “bad luck” for low-risk individuals to have multiple claims. The

estimates for the level of risk aversion and its dispersion are almost the same as in the benchmark model (Table 6). The rest of the summary figures are different. The difference in the level of risk is not informative: the interpretation of λ is slightly different, given the change in the underlying distribution. The dispersion in λ is much higher. This is a direct result of the thin-tail assumption. The dispersion of λ is identified by the tails of the claim distribution at the aggregate level. With thinner tails imposed on the individual distribution, more heterogeneity is needed to match the observed tails at the aggregate. Finally, the correlation coefficient in this specification changes signs and is now negative and close to -1 . This is consistent with our discussion in Section 3.3 that emphasizes that the identification of the correlation coefficient is closely tied to the structural assumptions. Once these are changed, the estimated correlation coefficient would change, too. Most importantly, however, the mean and dispersion of risk aversion are stable.

4.4.3 The distribution of absolute risk aversion

In the benchmark model, we assume that the coefficient of absolute risk aversion is lognormally distributed across individuals. Since only few previous studies focused on heterogeneity in risk preferences, there is not much existing evidence regarding the distribution of risk preferences. The only evidence we are aware of is the experimental results presented by Steffen Andersen et al. (2005), which show a skewed distribution with a fat right tail, which is qualitatively consistent with the lognormal distribution we assume.

An additional advantage of the normality assumption is computational. It 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. One may be concerned about the sensitivity of the results to this distributional assumption. For example, it may drive the result that the median level of risk aversion is much lower than the mean.

Incorporating alternative distributional assumptions significantly complicates and slows the estimation procedure.⁴⁴ As an alternative, we develop a procedure that, we believe, provides some guidance as to the sensitivity of the results to this distributional assumption. The disadvantage of the procedure is that it cannot account for adverse selection. Since we found that adverse selection is not that important, this exercise is informative. The exercise conveys that the main qualitative results are not driven by the (fat) tails of the lognormal distribution we impose. Rather, they are driven by the high fraction of individuals who chose low deductible despite being of low risk. The model implies that such individuals must have a fairly high level of risk aversion.

The exercise uses a Gibbs sampler to estimate the lognormal distribution of λ_i , given the covariates and the observed number of claims. In each iteration of this Gibbs sampler, conditional on the data and the most recent draws for the individual λ_i 's, we compute lower bounds for the level and dispersion of risk aversion (in addition to generating the draws for r_i from the conditional

⁴⁴A simple, discrete-type distribution of risk aversion is not well identified with the given data. This is due to the nature of the exercise. The distribution is identified by the binary deductible choice and by the “local” variation in the pricing menu. This variation is not enough to accurately pin down the discrete types.

lognormal distribution). To determine the lower bound for the level of risk aversion, we compute:

$$r_i(\lambda_i) = \begin{cases} 0 & \text{if } choice_i = 0 \\ \max \left[0, (\bar{d}_i)^{-1} \left(\frac{\Delta p_i}{\lambda_i \Delta d_i} - 1 \right) \right] & \text{if } choice_i = 1 \end{cases} \quad (18)$$

Namely, we make r_i as low as possible, given the assumptions of the model. We assume that individuals who chose a regular deductible are risk neutral, while individuals who chose a low deductible are just indifferent between the two deductible levels (unless λ_i is high enough, in which case they are also assumed to be risk neutral). We then compute the average over the r_i 's. To compute the lower bound for the dispersion, we search for values of r_i 's that are consistent with the assumptions of the model and the observed data, and that minimize the variance of r_i . This turns out to be a simple search procedure that is linear in the number of individuals.

The lower bound of the level of risk aversion that we obtain from this exercise is $3.68 \cdot 10^{-4} NIS^{-1}$ (with standard deviation of $6.38 \cdot 10^{-6}$). This level is five times lower than the analogous point estimate reported in Table 4. This translates into indifference about participating in a fifty-fifty lottery of gaining 100 dollars and losing 88.5. Thus, it still suggests a significant level of risk aversion for lotteries of 100 dollar magnitude. Similarly, the conversion to relative risk aversion, as in Table 5, implies a relative risk aversion coefficient of 18.83. The result for the lower bound of the dispersion is $1.89 \cdot 10^{-3}$ (with standard deviation of $8.16 \cdot 10^{-5}$), which is ten times lower than the benchmark estimate. Thus, the coefficient of variation for absolute risk aversion only declines by two, suggesting a high degree of heterogeneity. Finally, one should note that these bounds involve extreme assumptions, and are computed separately, and therefore cannot be binding at the same time. Thus, the ‘‘correct’’ estimates for both quantities are likely to be higher, closer to the results we report for the benchmark model.

4.4.4 Incomplete information about risk

Throughout, we assume that individuals have perfect information about their individual-specific objective risk type, λ_i . This is a stronger assumption than we need. Because expected utility is linear in probabilities, it suffices that individuals' *expected* risk rate is equal to their objective risk rate, i.e., $\hat{\lambda}_i \equiv E(\tilde{\lambda}_i | I_i) = \lambda_i$, where $\tilde{\lambda}_i$ is a random variable representing individual i 's perceived risk rate and I_i is individual i 's information at the time of the coverage choice. Namely, individuals could be uncertain about their risk type, but their point estimate has to be correct.

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 $\hat{\lambda}_i = \lambda_i + \epsilon_i$ 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 a relatively small effect of 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 an individual's estimate of his risk type is some weighted average between his true risk type and the average risk type of individuals who are similar (on observables) to him. The conclusion may go in the other way 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 plausible.

As a test for the sensitivity of the results to the complete information assumption, we perform two exercises. The first exercise estimates the benchmark model separately for experienced drivers and inexperienced drivers, where we define an individual to be an experienced driver if he has ten years or more of driving experience. This is a conservative definition compared to those used in the literature (Chiappori and Salanie, 2000; Cohen, 2005). The underlying assumption, consistent with evidence provided in Cohen (2005), is that drivers learn about their own risk types as they gain more driving experience, but that most of this learning occurs within the first ten years of driving. Thus, our complete information model is a more appealing description of experienced drivers. The results for experienced drivers are almost identical to the results from the benchmark specification (Table 6), suggesting that the main results are not driven by the fraction of individuals who are less likely to know their risk types. The results for inexperienced drivers show a similar pattern and magnitude. Consistent with the learning hypothesis, however, they show a larger dispersion in risk preferences, which may be due to incomplete information, and, therefore, to more random choices of deductible levels.

The second exercise we perform is a more structural version of the learning story, which allows even experienced drivers to have incomplete information regarding their risk types. We do so by estimating a different specification of the model that assumes that individuals are Bayesian and update their information about their own (stable) risk types over time, using only information about the number of claims they make each year. While we do not observe complete claims histories of individuals, we can simulate such histories and integrate over these simulations. Thus, individuals' information would be related to their true type, and would be more precise with longer driving histories. The full modeling and estimation details of this specification are provided in the appendix. We view this model as an extreme version of incomplete information, as there are many other sources through which individuals may learn about their own types, and thereby have better information about their types than what we estimate them to have. While the results imply that the level of risk aversion and heterogeneity are lower than the benchmark estimates (Table 6), the order of magnitude and qualitative pattern are quite similar, suggesting that the main qualitative findings are robust to the information structure.

4.4.5 Sample selection

There are two different ways in which one can think about the contract selection process. One possibility is that individuals first select an insurer based on advertisement, word-of-mouth, or access to an agent. Then individuals select the insurance contract from among several contracts the selected insurer offers. This selection process is consistent with the way the company executives view their business.⁴⁵ Another possibility is that individuals first collect information about all available insurance contracts and then choose their most preferred one. According to the industry's

⁴⁵As discussed in Section 2, the company was the first direct insurance provider and it offered significantly lower premiums than those offered by competitors due to significant cost advantage. In Section 2, we also discuss the literature, which emphasizes that choice of direct insurers is driven primarily by non-monetary "amenities."

conventional wisdom, individuals do not shop much across insurers and therefore this selection process seems less important.

The unique characteristics of direct insurers may attract individuals who are more likely to experiment with new ways to do business, and may be therefore less risk averse than the general population. In Table 7 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 slightly younger, more educated, more likely to be 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 suggests that the level of risk aversion that we find may be viewed as a lower bound on the level of risk aversion in the general Israeli 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 with 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. The results are still meaningful for two reasons. First, 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 informative.

To assess the magnitude of sample selection in driving the main results, we perform several tests. First, we estimate the benchmark model separately for the first two years of the company's operation, during which the company was the only company to sell insurance directly, and for the remaining three years, after additional direct insurers entered the market. Second, we estimate the benchmark model separately for individuals who were referred to the company by word-of-mouth and for those who heard about the company through advertisement (primarily on television). The latter may have searched more and might be more price sensitive. Third, we estimate the benchmark model separately for individuals who renewed their policy with the company and for those who did not renew. It seems likely that individuals who did not renew are less selected, as most of them switch back to regular insurers, who insure the majority of the population. Fourth, we estimate the model for the second deductible choice made by those individuals who renew. It could be argued that switching to other companies is costly, so outside options are not as important in driving deductible choices for those who renew. All these results are summarized in Table 6 and show some important differences between the two groups, within each pair of subsamples, probably reflecting selection. For all groups, however, the qualitative pattern of the results, and the order of magnitude and dispersion of risk aversion, are similar to those of the full sample. This is suggestive that correcting for sample selection is unlikely to change the qualitative results.

4.5 Caveats

4.5.1 Moral hazard

Throughout our analysis, we abstract from moral hazard, i.e., we assume 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. First, individuals with less coverage may take greater precaution and drive more carefully, thereby reducing their claim 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 (the literature often refers to this second effect as “ex post moral hazard”).

To the extent that moral hazard exists, abstracting from it 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. Similarly, ex post moral hazard would reduce the value of a low deductible, as, in the event of a claim, the gain from a lower deductible would sometimes be less than the difference between the two deductible levels. Both these effects will make a low deductible less attractive, requiring individuals to be even more risk averse than we estimate them to be in order to purchase more coverage. Below we discuss why, in our view, abstracting from moral hazard is a reasonable approximation in this setting.

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. However, when all choices include a deductible, and deductibles are similar in their magnitude, it seems less likely that driving or care behavior will be affected.⁴⁶ Finally, to separately identify moral hazard one would need another dimension of the data over which risk types remain fixed but coverage choices exogenously vary (Israel, 2004; Jaap H. Abbring et al., forthcoming).

We rely on the data to justify why we abstract from the second potential effect, that of ex post moral hazard. Figure 3 presents data on the claim amounts and shows that about 99 percent of the claims filed by policyholders with low-deductible policies were for amounts greater than the higher deductible level. If the distribution of amounts of potential claims does not vary with the deductible choice and if individuals file a claim for any loss that exceeds their deductible level, this suggests that 99 percent of the claims would have been filed under either deductible choice, making the assumption to abstract from moral hazard not very restrictive.

Individuals, however, may choose not to file a claim even when the claim amount exceeds the deductible level. This may happen due to experience rating, which increases future insurance premiums. 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

⁴⁶This assumption is consistent with Alma Cohen and Liran Einav (2003), who find no evidence for behavioral response to changes in seat belt laws. The following anecdotal observation also supports this. 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 driving or care behavior cannot depend on the deductible level.

with the company, we can assess how big the dynamic effects are. These data show that the price effect of a claim lasts for three years and is highest when an individual files his second claim within a year. In such a case, he would face about a 20 percent increase in his 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 about twice the regular deductible amount, so an upper bound for the dynamic costs is 70 percent of the regular deductible. In most cases, the actual dynamic costs are much lower than this upper bound (the dynamic costs of, say, the first claim within a year are close to zero). In addition, an individual can opt out of the contract and switch to a different insurance provider. This is likely to reduce his dynamic costs because in Israel, unlike in the U.S. and 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 premiums 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 (Cohen, 2003).⁴⁷ Using 70 percent as a conservative upper bound, Figure 3 shows that about 93 percent of those claims filed by individuals with a low deductible would have also been filed with a regular deductible. While this is not negligible, it applies for only a small fraction of the individuals. For the vast majority of them, the 99 percent discussed above is a more appropriate benchmark. 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.⁴⁸

4.5.2 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 incurred, 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. 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 \bar{d} . In particular, equation (7) will change to

$$r \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d} + c} \quad (19)$$

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 we present later. 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

⁴⁷New customers may voluntarily report their claims 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 0.2453 in our sample population. Our data on claims history, as voluntarily disclosed by the same individuals, suggest a claim rate of 0.0604, which is four times lower.

⁴⁸We would not be as comfortable with this statement for 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.

of $r\frac{\bar{d}+c}{\bar{d}}$, 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 this transaction cost c compared to the average deductible \bar{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 \bar{d} .

Data about such transaction costs are, of course, not available. The following back-of-the-envelope exercise may provide some guidance as to the magnitude of such costs. We collected data from the Israeli police about the annual numbers of accidents,⁴⁹ accidents with fatalities, and accidents with severe injuries in Israel for the years 1996-1999. We then divided these numbers by our estimate of the total number of auto insurance claims in Israel (161,859).⁵⁰ We obtain that 15.9 percent of the claims involve reported accidents, 2.2 percent involve accidents with severe injuries, and 0.3 percent involve fatal accidents. Thus, the majority of claims reflect small unreported accidents, perhaps suggesting that these additional costs of a claim are often not too large.

4.5.3 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 that facilitates comparison to previous studies. We 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 that are all in the range of 0.10 – 0.35. 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 2. 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 (Daniel Kahneman and Amos 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 arise if we try to apply decision weights (Amos Tversky and Peter Wakker, 1995) or measures of over-confidence with respect to driving ability.

⁴⁹ An accident is counted in this measure if it was reported to the police and a police officer arrived at the scene.

⁵⁰ Estimated by taking the number of active policies at the end of our sample (45,871), dividing it by our best guess for the share of the market the company had at the time (7 percent), and multiplying it by the estimated claim rate in our data, as computed in Table 2 (0.245).

4.6 Implications for profits and pricing

We now look at how firm profits vary with alternative pricing schemes. This exercise is interesting for several reasons. First, although we do not use supply-side information for estimation, it shows how one may incorporate such information in estimation. Second, it illustrates the conceptual trade-off faced by a monopolist that operates in a market with adverse selection. Although the conceptual trade-off between higher demand and worse selection is well known and has been extensively analyzed in the theoretical literature, quantifying it is important to understand its empirical relevance. Finally, we argued earlier that unobserved heterogeneity in risk aversion seems more important than unobserved heterogeneity in risk. The current exercise shows that this conclusion also translates to pricing and profits.

Throughout this section, we hold fixed the distribution of risk and risk aversion the firm faces. Since we have little information about the determinants of overall demand faced by the firm, and hence how restrictive it is to hold the distribution fixed, we make the simplifying assumption that individuals make their choices sequentially. They first choose the insurance provider by only observing the price of the regular deductible. Once they decide to buy a policy from the insurer, they choose the deductible level. This seems a reasonable approximation, 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.” Consistent with this assumption, we assume that the regular premium and deductible are dictated by competitive conditions, and we focus on the choice of the low deductible and its associated premium.

From the company’s standpoint, each individual can be represented by a random draw of (λ_i, r_i) from the conditional (on observables) distribution of risk and risk aversion:

$$\begin{pmatrix} \ln \lambda_i \\ \ln r_i \end{pmatrix} \sim N \left(\begin{pmatrix} x_i' \widehat{\beta} \\ x_i' \widehat{\gamma} \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_r \\ \rho \sigma_\lambda \sigma_r & \sigma_r^2 \end{pmatrix} \right) \quad (20)$$

When analyzing the optimal menu to offer such an individual, the company is assumed to be risk neutral and to maximize expected profits. Suppose the company only offered the regular deductible and premium (d_h, p_h) . Let the expected profits from this strategy be π_0 . Given our assumptions, we proceed by analyzing how the firm’s profits are affected by a choice of an additional low deductible option (d_l, p_l) , with $d_l < d_h$ and $p_l > p_h$. 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:

$$\max_{\Delta d, \Delta p} \{ \pi_0 + \Pr(r_i > r_i^*(\lambda_i; \Delta d, \Delta p)) [\Delta p - \Delta d \cdot E(\lambda_i | r_i > r_i^*(\lambda_i; \Delta d, \Delta p))] \} \quad (21)$$

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 deductible), but to lower quantity (or probability of purchase), as fewer individuals elect to choose the low deductible. This effect enters the profit function through $D(\Delta d, \Delta p) \equiv \Pr(r_i > r_i^*(\lambda_i; \Delta d, \Delta p))$. 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 $\bar{\lambda}(\Delta d, \Delta p) \equiv E(\lambda_i | r_i > r_i^*(\lambda_i; \Delta d, \Delta p))$. Its magnitude and sign depend on the relative heterogeneity of λ_i and r_i and on the correlation between them. Since neither $D(\Delta d, \Delta p)$ nor $\bar{\lambda}(\Delta d, \Delta p)$ have a closed-form solution, we analyze this decision problem graphically, where $D(\Delta d, \Delta p)$ and $\bar{\lambda}(\Delta d, \Delta p)$ are numerically computed using simulations from the joint distribution.

We illustrate our analysis by using the mean individual in the data, whose expected $\ln \lambda_i$ and $\ln r_i$ are -1.54 and -11.81 , respectively (based on Table 4). Such an individual has to choose between a regular contract of $(p_h, d_h) = (3189, 1595)$ (in NIS) and a low contract of $(p_l, d_l) = (3381, 957)$, i.e., $(\Delta p, \Delta d) = (191, 638)$. Below, we also discuss three additional counterfactual cases. First, we consider a case with a negative correlation between risk and risk aversion (with the same magnitude, i.e., $\rho = -0.84$). Second, we consider cases when the company ignores unobserved heterogeneity in one of the dimensions, i.e., it views individuals as a draw from the estimated marginal distribution on one dimension, with the other dimension known and fixed at its estimated mean. We do this exercise for each dimension separately.

To get intuition for the different effects, Figure 2 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 2 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 4 presents the effects of pricing on profits by varying the low-deductible level, keeping the premium charged for it fixed at the observed price (of $\Delta p = 191$). It implies that the current low deductible benefit of 638 NIS results in additional annual profits of about 3.68 NIS per customer. This is about 0.37 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, the relative magnitude of this effect will be higher. Note, also, that the estimates imply that the current low-deductible level is suboptimal. By setting a smaller low-deductible benefit of $\Delta d = 355$ NIS (i.e., increasing the current low deductible by 283 NIS), *additional* profits can be increased to 6.59 NIS.⁵¹ Of course, the finding that the current pricing is suboptimal may also be due to a limitation of the model.

Consistent with the intuition discussed above, Figure 4 also shows that the incentive to increase prices (or lower the deductible) is higher with positive correlation. When the correlation between

⁵¹There is no reason, of course, to limit the choice of the company to only one additional deductible level. More degrees of freedom in choosing the pricing menu will lead to higher profits.

risk and risk aversion is negative, the optimal low deductible is lower. It also shows that ignoring either dimension of unobserved heterogeneity has an important effect on pricing decisions. However, while ignoring heterogeneity in risk does not change the qualitative pattern by much, ignoring heterogeneity in risk aversion completely changes the pricing analysis. In fact, given the estimated coefficients and the observed prices, if individuals only vary in risk, offering a low deductible does not increase profits, due to adverse selection.

Figure 5 breaks down the effect on profits by presenting the pricing effect on the demand for low deductible, $D(\Delta d, \Delta p)$, and on the composition effect, $\bar{\lambda}(\Delta d, \Delta p)$. The former is simply generated by the distribution of certainty equivalents implied by the joint distribution of λ_i and r_i (Landsberger and Meilijson, 1999). It is S-shaped due to the assumption of lognormal distribution. The shape of the composition effect is driven by the relative variance of λ_i and r_i and by the correlation coefficient. 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 choosing it first. The effect of the deductible level on the composition effect is dramatically different when the correlation between risk and risk aversion is negative. When this is the case, 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.

5 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 many other contexts, 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 for wealth after death (Liran Einav et al., 2006). Similarly, one could consider loan data and use down payment choices and default data to identify between heterogeneity in risk (default) and in liquidity (will Adams et al., 2006). Second, from an economic standpoint, we provide a new set of estimates for the degree and heterogeneity of (absolute) risk aversion, and its relationship with risk. We discuss these below.

While our estimates of risk aversion help to predict other related insurance decisions, 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, 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. On the other hand, classical theory suggests that each individual has one value function over her lifetime 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. To assess “closeness,” it is important to consider various factors over which contexts may differ. Such factors may include bet size, as well as informational and behavioral effects, such as default options, framing, and rarity of the events. Since we use claim data to identify risk type, we leave everything else to be interpreted as risk aversion. As an example, over-confidence will be captured by a lower level of estimated risk aversion, so if over-confidence is more important in auto insurance than in health insurance, then when extrapolated to health insurance, individuals may behave as if they are more risk averse than we estimate them to be.

With this caveat in mind, let us discuss our main four findings. First, we find large heterogeneity in risk preferences across individuals. This heterogeneity is important in various contexts: (i) it cautions against using pure cross-sectional variation to test expected utility theory; (ii) it may create strong selection of participants into particular markets: this may make participants in voluntary markets quite different in their risk preferences from those in markets in which participation is either mandatory or driven by other factors; and (iii) it may make the marginal individual quite different from the average one. Models in macroeconomics and finance, which often use a representative individual framework, may not be able to capture and account for such differences.

Our second set of findings concerns the way risk aversion relates to observable characteristics. Our finding that women are more risk averse than men has been documented in other settings. The finding that risk preferences exhibit a U-shaped pattern over the life cycle may be interesting to explore further in other contexts. Other findings suggest that the estimated coefficient of risk aversion increases with observables that are related to income and wealth. As we agree with the widely held belief of the decreasing absolute risk aversion property, our preferred interpretation for this finding is that wealth and income may be endogenous, generating the estimated cross-sectional relationship. 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 relationship between risk and preferences. Since risk is particular to the context in which it is measured, these findings may be sensitive to the market context, and may change once risk takes other forms. Even within the auto insurance market, it is important how risk is measured. Moreover, risk in the auto insurance market may be conceptually different from risk in other markets. In many markets risk is independent across individuals. In auto insurance, however, much of the risk depends on coordination among drivers, and therefore may be more related to relative, not absolute, characteristics. For this reason, our finding of positive correlation between risk and risk aversion can coexist with findings of negative correlation in other contexts (Israel, 2005; Finkelstein and McGarry, 2006). The positive correlation we find is also

consistent with the fact that the bivariate probit test in our data provides evidence for adverse selection (Cohen, 2005), while similar reduced form tests in other contexts do not.

Finally, we find that unobserved heterogeneity in risk preferences is more important than heterogeneity in risk. This may 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 about preferences.⁵² We illustrate the empirical importance 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. 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 (Jean-Charles Rochet and Lars A. Stole, 2002), or assumes that the number of instruments available to the monopolist is not smaller than the dimension of unobserved heterogeneity (Steven Matthews and John Moore, 1987; Richard Arnott and Joseph E. Stiglitz, 1988).⁵³ Mark Armstrong (1999) 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 restrictions in similar contexts. Our counterfactual analysis is a very preliminary start in this direction.

⁵²This choice of data collection efforts may be justified if it is easier for such firms to price discriminate based on risk, but harder to price discriminate based on preferences. A common belief is that, without cost-based (i.e., risk-based) justification for prices, price discrimination may lead to consumer backlash.

⁵³See also, Smart (2000), Villeneuve (2003), and Jullien et al. (forthcoming) for related theoretical results.

References

- [1] **Abbring, Jaap H., Pierre-Andre Chiappori, James Heckman, and Jean Pinquet.** Forthcoming. “Testing for Moral Hazard on Dynamic Insurance Data.” *Journal of the European Economic Association Papers and Proceedings*.
- [2] **Abbring, Jaap H., Pierre-Andre Chiappori, and Jean Pinquet.** 2003. “Moral Hazard and Dynamic Insurance Data.” *Journal of the European Economic Association*, 1(4): 767-820.
- [3] **Adams, Will, Liran Einav, and Jonathan D. Levin.** 2006. “Screening in Consumer Credit Markets.” Unpublished.
- [4] **Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson.** 2006. “Financial Mistakes Over the Life Cycle.” Unpublished.
- [5] **Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutstrom.** 2005. “Preference Heterogeneity in Experiments: Comparing the Field and Lab.” Centre for Economic and Business Research Discussion Paper 2005-03.
- [6] **Armstrong, Mark.** 1999. “Optimal Regulation with Unknown Demand and Cost Functions.” *Journal of Economic Theory*, 84(2): 196-215.
- [7] **Arnott, Richard, and Joseph E. Stiglitz.** 1988. “Randomization with Asymmetric Information.” *Rand Journal of Economics*, 29(3): 344-362.
- [8] **Auld, M. Christopher, J. C. Herbert Emery, Daniel V. Gordon, and Douglas McClintock.** 2001. “The Efficacy of Construction Site Safety Inspections.” *Journal of Labor Economics*, 19(4): 900-921.
- [9] **Barberis, Nicholas, Ming Huang, and Richard H. Thaler.** 2006. “Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing.” *American Economic Review*, 96(4): 1069-1090.
- [10] **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.
- [11] **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(474): 821-848.
- [12] **Bombardini, Matilde, and Francesco Trebbi.** 2005. “Risk Aversion and Expected Utility Theory: A Field Experiment with Large and Small Stakes,” <http://www.econ.ubc.ca/mbombardini/index.htm>.
- [13] **Bortkiewicz, Ladislaus.** 1898. *Das Gesetz der Kleinen Zahlen* (The Law of Small Numbers). Leipzig, Germany: Teubner.

- [14] **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.
- [15] **Cawley, John, and Tomas Philipson.** 1999. "An Empirical Examination of Information Barriers to Trade in Insurance." *American Economic Review*, 89(4): 827-846.
- [16] **Chetty, Raj.** Forthcoming. "A New Method of Estimating Risk Aversion." *American Economic Review*.
- [17] **Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie.** Forthcoming. "Asymmetric Information in Insurance: General Testable Implications." *Rand Journal of Economics*.
- [18] **Chiappori, Pierre-Andre, and Bernard Salanie.** 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108(1): 56-78.
- [19] **Chiappori, Pierre-Andre, and Bernard Salanie.** 2006. "The Microeconomics of Insurance: An Empirical Perspective." Paper presented at the annual meeting of the American Economic Association, Boston.
- [20] **Choi, Syngjoo, Douglas Gale, Ray Fisman, and Shachar Kariv.** 2006. "Substantive and Procedural Rationality in Decisions under Uncertainty." <http://socrates.berkeley.edu/~kariv>.
- [21] **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.
- [22] **Cohen, Alma.** 2003. "Profits and Market Power in Repeat Contracting: Evidence from the Insurance Market." Unpublished.
- [23] **Cohen, Alma.** 2005. "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market." *Review of Economics and Statistics*, 87(2): 197-207.
- [24] **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.
- [25] **Cooper, Russell, and Beth Hayes.** 1987. "Multi-period Insurance Contracts." *International Journal of Industrial Organization*, 5(2): 211-231.
- [26] **Cummins, J. David, and Jack L. Van Derhei.** 1979. "A Note on the Relative Efficiency of Property-Liability Insurance Distribution Systems." *Bell Journal of Economics*, 10(2): 709-719.
- [27] **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.

- [28] **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.
- [29] **De Meza, David, and David C. Webb.** 2001. “Advantageous Selection in Insurance Markets.” *Rand Journal of Economics*, 32(2): 249-262.
- [30] **Devroye, Luc.** 1986. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
- [31] **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.
- [32] **Dionne, Georges, and Pierre Lasserre.** 1985. “Adverse Selection, Repeated Insurance Contracts, and Announcement Strategy.” *Review of Economic Studies*, 52(4): 719-723.
- [33] **Dionne, Georges, and Charles Vanasse.** 1992. “Automobile Insurance Ratemaking in the Presence of Asymmetrical Information.” *Journal of Applied Econometrics*, 7(2): 149-165.
- [34] **Donkers, Bas, Bertrand Melenberg, and Arthur van Soest.** 2001. “Estimating Risk Attitudes Using Lotteries: A Large Sample Approach.” *Journal of Risk and Uncertainty*, 22(2): 165-195.
- [35] **Dreze, Jacques H.** 1981. “Inferring Risk Tolerance from Deductibles in Insurance Contracts.” *Geneva Papers on Risk and Insurance*, 20(July): 48-52.
- [36] **Einav, Liran, Amy Finkelstein, and Paul Schrimpf.** 2006. “The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market.” <http://www.stanford.edu/~leinav>.
- [37] **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.
- [38] **Finkelstein, Amy, and Kathleen McGarry.** 2006. “Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market.” *American Economic Review*, 96(4): 938-958.
- [39] **Finkelstein, Amy, and James Poterba.** 2004. “Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market.” *Journal of Political Economy*, 112(1): 183-208.
- [40] **Friend, Irwin, and Marshall E. Blume.** 1975. “The Demand for Risky Assets.” *American Economic Review*, 65(5): 900-922.
- [41] **Gertner, Robert.** 1993. “Game Shows and Economic Behavior: Risk-Taking on ‘Card Sharks’.” *Quarterly Journal of Economics*, 108(2): 507-521.

- [42] **Geweke, John.** 1991. "Efficient Simulation from the Multivariate Normal and Student-t Distributions subject to Linear Constraints." In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E. M. Keramidas, 571-578. Interface Foundation of North America.
- [43] **Hartog, Joop, Ada Ferrer-i-Carbonell, and Nicole Jonker.** 2002. "Linking Measured Risk Aversion to Individual Characteristics." *Kyklos*, 55(1): 3-26.
- [44] **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.
- [45] **Holt, Charles A., and Susan K. Laury.** 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5): 1644-1655.
- [46] **Israel, Mark.** 2004. "Do We Drive More Safely When Accidents Are More Expensive? Identifying Moral Hazard from Experience Rating Schemes." <http://www.kellogg.northwestern.edu/faculty/israel/htm/research.html>.
- [47] **Israel, Mark.** 2005. "Where Is All the Hidden Information Hiding? Evidence from Automobile Insurance Panel Data." <http://www.kellogg.northwestern.edu/faculty/israel/htm/research.html>.
- [48] **Jullien, Bruno, and Bernard Salanie.** 2000. "Estimating Preferences under Risk: The Case of Racetrack Bettors." *Journal of Political Economy*, 108(3): 503-530.
- [49] **Jullien, Bruno, Bernard Salanie, and Francois Salanie.** Forthcoming. "Screening Risk-Averse Agents under Moral Hazard: Single-Crossing and the CARA Case." *Economic Theory*.
- [50] **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.
- [51] **Kahneman, Daniel, and Amos Tversky.** 1979. "Prospect Theory: An Analysis of Decisions under Risk." *Econometrica*, 47(2): 263-292.
- [52] **Kimball, Miles S., and N. Gregory Mankiw.** 1989. "Precautionary Saving and the Timing of Taxes." *Journal of Political Economy*, 97(4): 863-879.
- [53] **Kocherlakota, Narayana R.** 1996. "The Equity Premium: It's Still a Puzzle." *Journal of Economic Literature*, 34(1): 42-71.
- [54] **Kunreuther, Howard, and Mark V. Pauly.** 1985. "Market Equilibrium with Private Knowledge: An Insurance Example." *Journal of Public Economics*, 26(3): 269-288.
- [55] **Landsberger, Michael, and Isaac Meilijson.** 1999. "A General Model of Insurance under Adverse Selection." *Economic Theory*, 14(2): 331-352.

- [56] **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.
- [57] **Metrick, Andrew.** 1995. "A Natural Experiment in 'Jeopardy!'" *American Economic Review*, 85(1): 240-253.
- [58] **Michener, Ron, and Carla Tighe.** 1992. "A Poisson Regression Model of Highway Fatalities." *American Economic Review Papers and Proceedings*, 82(2): 452-456.
- [59] **Palacios-Huerta, Ignacio, Roberto Serrano, and Oscar Volij.** 2006. "Rejecting Small Gambles Under Expected Utility." *Economics Letters*, 91(2): 250-259.
- [60] **Puelz, Robert, and Arthur Snow.** 1994. "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market." *Journal of Political Economy*, 102(2): 236-257.
- [61] **Rabin, Matthew.** 2000. "Risk Aversion and Expected Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281-1292.
- [62] **Rabin, Matthew, and Richard H. Thaler.** 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives*, 15(1): 219-232.
- [63] **Rochet, Jean-Charles, and Lars A. Stole.** 2002. "Nonlinear Pricing with Random Participation." *Review of Economic Studies*, 69(1): 277-311.
- [64] **Rose, Nancy L.** 1990. "Profitability and Product Quality: Economic Determinants of Airline Safety Performance." *Journal of Political Economy*, 98(5): 944-964.
- [65] **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.
- [66] **Rubinstein, Ariel.** 2001. "Comments on the Risk and Time Preferences in Economics." <http://arielrubinstein.tau.ac.il/vitae.html>.
- [67] **Saha, Atanu.** 1997. "Risk Preference Estimation in the Nonlinear Mean Standard Deviation Approach." *Economic Inquiry*, 35(4): 770-782.
- [68] **Smart, Michael.** 2000. "Competitive Insurance Markets with Two Unobservables." *International Economic Review*, 41(1): 153-169.
- [69] **Smith, Vernon L., and James M. Walker.** 1993. "Rewards, Experience, and Decision Costs in First-Price Auctions." *Economic Inquiry*, 31(2): 237-244.
- [70] **Sydnor, Justin.** 2006. "Sweating the Small Stuff: The Demand for Low Deductibles in Homeowners Insurance." <https://webfiles.berkeley.edu/~jrsydnor>.

- [71] **Tanner, Martin A., and Wing Hung Wong.** 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association*, 82(398): 528-549.
- [72] **Tversky, Amos, and Peter Wakker.** 1995. "Risk Attitudes and Decision Weights." *Econometrica*, 63(6): 1255-1280.
- [73] **Villeneuve, Bertrand.** 2003. "Concurrence et Antisélection Multidimensionnelle en Assurance," *Annales d'Economie et Statistique*, 69(January-March): 119-142.
- [74] **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.
- [75] **Watt, Richard.** 2002. "Defending Expected Utility Theory." *Journal of Economic Perspectives*, 16(2): 227-230.

Appendix

A Estimation algorithm

A.1 Gibbs sampler for the benchmark specification

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:

$$\ln \lambda_i = x_i' \beta + \varepsilon_i \quad (22)$$

$$\ln r_i = x_i' \gamma + v_i \quad (23)$$

$$choice_i = \begin{cases} 1 & \text{if } r_i > r_i^*(\lambda) \\ 0 & \text{if } r_i < r_i^*(\lambda) \end{cases} \quad (24)$$

$$claims_i \sim Poisson(\lambda_i t_i) \quad (25)$$

$$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \overset{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_r \\ \rho \sigma_\lambda \sigma_r & \sigma_r^2 \end{bmatrix} \right) \quad (26)$$

$$\text{Let } \delta \equiv \begin{bmatrix} \beta \\ \gamma \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_r \\ \rho \sigma_\lambda \sigma_r & \sigma_r^2 \end{bmatrix}, X \equiv \begin{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{bmatrix}, y \equiv \begin{bmatrix} \lambda \\ r \end{bmatrix}, \text{ and } u_i \equiv$$

$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix}$. 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 $\{u_i\}_{i=1}^n$:

$$\{u_i\}_{i=1}^n | \Sigma \overset{iid}{\sim} N(0, \Sigma) \quad (27)$$

$$\Sigma^{-1} \sim Wishart_2(a, Q) \quad (28)$$

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

$$\Sigma^{-1} | \delta, \{u_i\}_{i=1}^n \sim Wishart_2 \left(a + n - k, \left(Q^{-1} + \sum_i u_i u_i' \right)^{-1} \right) \quad (29)$$

and

$$\delta|\Sigma, \{u_i\}_{i=1}^n \sim N((X'X)^{-1}(X'y), \Sigma^{-1} \otimes (X'X)^{-1}) \quad (30)$$

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

The part of the Gibbs sampler that 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 $v_i = \ln r_i - x_i'\gamma$, so we can instead focus on sampling from the posterior distribution of λ_i and r_i . These posterior distributions are:

$$\Pr(r_i|\gamma, \beta, \Sigma, \lambda_i, data) \propto \begin{cases} \phi\left[\ln r_i, x_i'\gamma + \rho\frac{\sigma_r}{\sigma_\lambda}(\ln \lambda_i - x_i'\beta), \sqrt{\sigma_r^2(1-\rho^2)}\right] & \text{if } choice_i = I(r_i < r_i^*(\lambda)) \\ 0 & \text{if } choice_i \neq I(r_i < r_i^*(\lambda)) \end{cases} \quad (31)$$

and

$$\Pr(\lambda_i|\gamma, \beta, \Sigma, r_i, data) \propto \begin{cases} p(\lambda_i, claims_i, t_i) \phi\left[\ln \lambda_i, x_i'\beta + \rho\frac{\sigma_\lambda}{\sigma_r}(\ln r_i - x_i'\gamma), \sqrt{\sigma_\lambda^2(1-\rho^2)}\right] & \text{if } choice_i = I(r_i < r_i^*(\lambda)) \\ 0 & \text{if } choice_i \neq I(r_i < r_i^*(\lambda)) \end{cases} \quad (32)$$

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

The posterior for $\ln r_i$ is a truncated normal, for which we use a simple ‘‘invert c.d.f.’’ sampling (Devroye, 1986).⁵⁴ The posterior for $\ln \lambda_i$ is less standard. 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 a 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 $\Pr(\lambda_i, u_i^1, u_i^2) = b_0(\lambda_i)b_1(\lambda_i)b_2(\lambda_i)\frac{I(0 \leq u_i^1 \leq b_1(\lambda_i))}{b_1(\lambda_i)}\frac{I(0 \leq u_i^2 \leq b_2(\lambda_i))}{b_2(\lambda_i)} = b_0(\lambda_i)I(0 \leq u_i^1 \leq b_1(\lambda_i))I(0 \leq u_i^2 \leq b_2(\lambda_i))$. 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_i^1 and u_i^2 , this just means that $b_1^{-1}(u_i^1) = \frac{\ln u_i^1}{claims_i}$ is a lower bound for $\ln \lambda_i$ (for $claims_i > 0$) and that $b_2^{-1}(u_i^2) = \ln(-\ln u_i^2) - \ln t_i$ 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 .

⁵⁴Let $F(x)$ be the cumulative distribution function. The ‘‘invert c.d.f.’’ sampling draws from this distribution by drawing u from a uniform distribution on $[0, 1]$ and computing $F^{-1}(u)$. In principle, one can use the sampling procedure suggested by John Geweke (1991), which avoids computing $F^{-1}(\cdot)$ and therefore is more efficient. It was easier, however, to vectorize the algorithm using Devroye (1986). The vectorization entails enormous computational benefits when coded in Matlab.

A.2 Gibbs sampler for the learning model

In the end of Section 4.4, we add to the model incomplete information of individuals about their own types. Individuals’ types are fixed over their lifetime, and individuals are Bayesian and update 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. Before individuals obtain their driving license, they believe that their risk type is a random draw from the observed population of drivers in the data. Individuals are assumed to have a prior that follows a *Gamma*(α, β) distribution,⁵⁵ where α and β are estimated from the data. Individuals’ posterior mean is then 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 $\hat{\lambda}_i$ denote the posterior mean. The assumptions imply that $\hat{\lambda}_i(l_i + \frac{1}{\beta}) - \alpha$ is distributed *Poisson*($\lambda_i l_i$). The rest of the model is as before, with $\hat{\lambda}_i$ used instead of λ_i to explain the coverage choice. Thus, to implement it within the Gibbs sampler, we augment $\hat{\lambda}_i$ as well. The conditional distribution for $r_i | \lambda_i, \hat{\lambda}_i$ is as before, with $\hat{\lambda}_i$ (rather than λ_i) affecting the truncation point. The conditional distribution for $\hat{\lambda}_i | \lambda_i, r_i$ is a linear transformation of a truncated Poisson, with the truncation point being the same as the one used above for λ_i . Finally, the conditional distribution for $\lambda_i | \hat{\lambda}_i, r_i$ is 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, $\hat{\lambda}_i$ provides information on λ_i . In particular, it can be written as:

$$\begin{aligned} & \Pr \left(\lambda_i | \gamma, \beta, \Sigma, \hat{\lambda}_i, r_i, data \right) \propto \\ & \propto p(\lambda_i, \hat{\lambda}_i(l_i + \frac{1}{\beta}) - \alpha, l_i) p(\lambda_i, claims_i, t_i) \phi \left[\ln \lambda_i, x'_i \beta + \rho \frac{\sigma_\lambda}{\sigma_r} (\ln r_i - x'_i \gamma), \sqrt{\sigma_\lambda^2 (1 - \rho^2)} \right] \end{aligned} \quad (33)$$

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 that 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 that is equal to 1 if the individual was not born in Israel.
- (Car) value - Current estimated “Blue Book” value of the car.
- Car age - The number of years the car has been in use.
- Commercial car - A dummy variable that is equal to 1 if the car is defined as a commercial vehicle (e.g., pickup truck).

⁵⁵Note that the gamma assumption is similar, but not identical, to the lognormal 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.

- Engine size - The volume of the engine in cubic centimeters (cc). This is a measure of size and power. For modern cars, 1 unit of horsepower is roughly equal to 15-17 cc, depending on several other variables, such as weight.
- License years - Number of years since the individual obtained driving license.
- Good driver - A dummy variable that 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.
- Any driver - A dummy variable that 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 policyholder (and sometimes his/her spouse) drives the car.
- Secondary car - A dummy variable that is equal to 1 if the car is not the main car in the household.
- Business use - A dummy variable that is equal to 1 if the policyholder uses the car for business.
- Estimated mileage - Predicted annual mileage (in kilometers) by the policyholder. The company does not use this variable for pricing, as it is believed to be unreliable.
- History - The number of years (up to the required 3) prior to the starting date of the policy for which the policyholder reports his/her past claims history.
- Claims history - The number of claims per year for the policyholder over the 3 (or less) years prior to the starting date of the policy.
- Young driver - A dummy variable that is equal to 1 if the policy covers drivers who are below the age of 25. In such cases, the policyholder has to separately report the details of the young driver (which may be the policyholder or someone else).
- Company year - Year dummies that 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 on.

Table 1: Summary statistics – covariates

Variable		Mean	Std. Dev.	Min	Max	
Demographics:	Age	41.137	12.37	18.06	89.43	
	Female	0.316	0.47	0	1	
	Family	Single	0.143	0.35	0	1
		Married	0.779	0.42	0	1
		Divorced	0.057	0.23	0	1
		Widower	0.020	0.14	0	1
		Refused to Say	0.001	0.04	0	1
	Education	Elementary	0.016	0.12	0	1
		High School	0.230	0.42	0	1
		Technical	0.053	0.22	0	1
		Academic	0.233	0.42	0	1
		No Response	0.468	0.50	0	1
	Emigrant	0.335	0.47	0	1	
	Car Attributes:	Value (current NIS) ¹	66,958	37,377	4,000	617,000
Car Age		3.952	2.87	0	14	
Commercial Car		0.083	0.28	0	1	
Engine Size (cc)		1,568	385	700	5,000	
Driving:	License Years	18.178	10.07	0	63	
	Good Driver	0.548	0.50	0	1	
	Any Driver	0.743	0.44	0	1	
	Secondary Car	0.151	0.36	0	1	
	Business Use	0.082	0.27	0	1	
	Estimated Mileage (km) ²	14,031	5,891	1,000	32,200	
	History Length	2.847	0.61	0	3	
	Claims History	0.060	0.15	0	2	
Young Driver:	Young	0.192	0.39	0	1	
	Gender	Male	0.113	0.32	0	1
		Female	0.079	0.27	0	1
	Age	17-19	0.029	0.17	0	1
		19-21	0.051	0.22	0	1
		21-24	0.089	0.29	0	1
		>24	0.022	0.15	0	1
	Experience	<1	0.042	0.20	0	1
		1-3	0.071	0.26	0	1
		>3	0.079	0.27	0	1
Company Year:	First Year	0.207	0.41	0	1	
	Second Year	0.225	0.42	0	1	
	Third Year	0.194	0.40	0	1	
	Fourth Year	0.178	0.38	0	1	
	Fifth Year	0.195	0.40	0	1	

The table is based on all 105,800 new customers in the data.

¹ 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.

² The estimated mileage is not reported by everyone. It is available for only 60,422 new customers.

Table 2A: Summary statistics – menus, choices, and outcomes

Variable			Obs	Mean	Std. Dev.	Min	Max	
Menu:	Deductible (current NIS) ¹	Low	105,800	875.48	121.01	374.92	1,039.11	
		Regular	105,800	1,452.99	197.79	624.86	1,715.43	
		High	105,800	2,608.02	352.91	1,124.75	3,087.78	
		Very High	105,800	3,763.05	508.53	1,624.64	4,460.13	
	Premium (current NIS) ¹	Low	105,800	3,380.57	914.04	1,324.71	19,239.62	
		Regular	105,800	3,189.22	862.3	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	
	<i>Δp/Δd</i>			105,800	0.328	0.06	0.3	1.8
	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 Duration (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 ²		All	105,800	0.245	0.66	0	198.82	
		Low	18,799	0.309	0.66	0	92.64	
		Regular	85,840	0.232	0.66	0	198.82	
		High	654	0.128	0.62	0	126.36	
		Very High	507	0.133	0.50	0	33.26	

¹ 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.

² The mean and standard deviation of the claims per year are weighted by the observed policy duration to adjust for variation in the exposure period. These are the maximum likelihood estimates of a simple Poisson model with no covariates.

Table 2B: Summary statistics - 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%)	80.34%
1	3,124 (23.9%)	9,867 (75.5%)	47 (0.4%)	35 (0.3%)	13,073 (100%)	16.96%
2	565 (30.8%)	1,261 (68.8%)	4 (0.2%)	2 (0.1%)	1,832 (100%)	2.38%
3	71 (31.4%)	154 (68.1%)	1 (0.4%)	0	226 (100%)	0.29%
4	6 (35.3%)	11 (64.7%)	0	0	17 (100%)	0.02%
5	1 (50.0%)	1 (50.0%)	0	0	2 (100%)	0.00%

Table 2B presents tabulation of choices and number of claims. For comparability, 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 in Table 2B 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: No heterogeneity in risk

Variable		Poisson Regression ¹			Probit Regressions ³				
		Dep. Var: Number of Claims			Dep. Var: 1 if Low Deductible				
		(1)			(2)			(3)	
		Coeff	Std. Err.	IRR ²	Coeff	Std. Err.	dP/dX	dP/dX	
Demographics:	Constant	-4.3535 [^]	0.2744	-	-22.4994 [^]	2.3562	-		
	Age	-0.0074	0.0052	0.9926	-0.1690 [^]	0.0464	-0.0034 [^]	-0.0042 [^]	
	Age ²	9.89·10 ⁻⁵	5.36·10 ⁻⁵	1.0001	0.0017 [^]	0.0005	3.39·10 ⁻⁵ [^]	4.55·10 ⁻⁵ [^]	
	Female	-0.0456 [^]	0.0161	0.9555	0.8003 [^]	0.1316	0.0165 [^]	0.0134 [^]	
	Family			omitted			omitted	omitted	
		Single							
		Married	-0.1115 [^]	0.0217	0.8945	0.6417 [^]	0.1950	0.0128 [^]	0.0045
		Divorced	0.0717 [^]	0.0346	1.0743	-0.3191	0.3127	-0.0064	0.0044
		Widower	0.0155	0.0527	1.0156	-0.3441	0.4573	-0.0069	-0.0025
		Other (NA)	0.1347	0.1755	1.1442	-2.9518	1.9836	-0.0511	-0.037
	Education	Elementary	-0.0522	0.0550	0.9491	0.2927	0.4278	0.0061	0.0016
		High School			omitted			omitted	omitted
		Technical	0.0373	0.0308	1.0380	0.5037 [^]	0.252	0.0105 [^]	0.0138 [^]
		Academic	-0.0745 [^]	0.0197	0.9282	0.5279 [^]	0.1546	0.0109 [^]	0.0053
	Other (NA)	0.0116	0.0184	1.0116	0.063	0.154	0.0013	0.0025	
	Emigrant	0.0210	0.0163	1.0213	-0.0497	0.1328	-0.0010	0.0003	
Car Attributes:	Log(Value)	0.1227 [^]	0.0281	1.1306	1.3285 [^]	0.2342	0.0271 [^]	0.0299 [^]	
	Car Age	0.0187 [^]	0.0042	1.0189	-0.1603 [^]	0.0357	-0.0033 [^]	-0.0017 [^]	
	Commercial Car	-0.1394 [^]	0.0326	0.8699	-0.9038 [^]	0.2781	-0.0177 [^]	-0.0294 [^]	
	Log(Engine Size)	0.2972 [^]	0.0459	1.3461	-0.6924	0.3952	-0.0141	0.0075	
Driving:	License Years	-0.0204 [^]	0.0034	0.9798	0.1043 [^]	0.0312	0.0021 [^]	0.0005	
	License Years ²	0.0002 [^]	6.77·10 ⁻⁵	1.0002	-0.0015 [^]	0.0005	-2.99·10 ⁻⁵ [^]	-1.67·10 ⁻⁵	
	Good Driver	-0.0176	0.0191	0.9825	-0.7207 [^]	0.1574	-0.0148 [^]	-0.0152 [^]	
	Any Driver	0.0564 [^]	0.0169	1.0580	1.0923 [^]	0.1321	0.0217 [^]	0.0258 [^]	
	Secondary Car	-0.0859 [^]	0.0209	0.9177	0.0038	0.1626	0.0001	-0.0070 [^]	
	Business Use	0.1852 [^]	0.0293	1.2034	-0.7951 [^]	0.2737	-0.0156	-0.0017	
	History Length	-0.0527 [^]	0.0110	0.9486	1.1218 [^]	0.1450	0.0228 [^]	0.0171 [^]	
	Claims History	0.6577 [^]	0.0390	1.9304	-0.4654	0.5460	-0.0095	0.0496 [^]	
Young Driver:	Young driver	0.5235 [^]	0.0361	1.6879	-2.8847 [^]	0.6305	-0.0524 [^]	-0.0012	
	Gender			omitted			omitted	omitted	
		Male							
		Female	-0.1475 [^]	0.0288	0.8629	1.7959 [^]	0.3135	0.0396 [^]	0.0195 [^]
	Age			omitted			omitted	omitted	
		17-19							
		19-21	0.0701	0.0532	1.0726	-0.7800	0.7509	-0.0153	-0.0147
		21-24	-0.0267	0.0574	0.9737	-0.5746	0.7773	-0.0114	-0.0156
	>24	-0.2082 [^]	0.0567	0.812	1.7869 [^]	0.7328	0.0397 [^]	0.0136	
Experience	<1			omitted			omitted	omitted	
	1-3	-0.2416 [^]	0.0458	0.7854	1.0175	0.6577	0.0217	-0.0009	
	>3	-0.2827 [^]	0.0532	0.7538	3.2513 [^]	0.7386	0.0762 [^]	0.0410 [^]	
Company Year:	First Year			omitted			omitted	omitted	
	Second Year	-0.0888 [^]	0.0198	0.9150	-4.4513 [^]	0.1506	-0.0792 [^]	-0.0859 [^]	
	Third Year	-0.0690 [^]	0.0222	0.9334	-8.5888 [^]	0.1820	-0.1303 [^]	-0.1367 [^]	
	Fourth Year	-0.1817 [^]	0.0232	0.8339	-11.8277 [^]	0.2102	-0.1616 [^]	-0.1734 [^]	
	Fifth Year	-0.5431 [^]	0.028	0.5810	-14.3206 [^]	0.2778	-0.1920 [^]	-0.2085 [^]	
σ				10.5586					
Obs	105,800			94,000			105,800		
Pseudo R ²	0.0162			0.1364			0.1296		
Log(Likelihood)	-57,745.9			-36.959.5			-43.086.9		

[^] Significant at the five-percent confidence level.

¹ Maximum likelihood estimates. Variation in exposure (policy duration) is accounted for.

² 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.

³ There are two separate probit regressions reported in this table. Column (2) relies on the deductible choice model and the lognormality assumption. As discussed in the text, by including an additional regressor, $\ln\left(\frac{\Delta p_i / (\hat{\lambda}(x_i) \Delta d_i) - 1}{d_i}\right)$ (with $\hat{\lambda}(x_i)$ predicted from column (1) above) and normalizing its coefficient to -1, we obtain a structural interpretation of this regression, with $\ln r_i$ as the dependent variable. Thus, the reported coefficients are comparable to those estimated for the benchmark model. However, one should be cautious in interpreting these coefficients. Unlike the benchmark model, this regression does not allow unobserved heterogeneity in risk and suffers from some selection bias because observations with high predicted risk rate are omitted (which is why the number of observations is 94,000 rather than the full sample of 105,800). For comparison, column (3) reports the marginal effects from a comparable probit regression that uses the full sample and does not control for pricing and predicted risk through the additional structural regressor. Column (3) does not have a structural interpretation, and its (unreported) coefficients cannot be compared to those estimated from the benchmark model.

Table 4: The benchmark model

Variable		Ln(λ) Equation	Ln(r) Equation	Additional Quantities	
Demographics:	Constant	-1.5406 [^] (0.0073)	-11.8118 [^] (0.1032)	Var-Covar Matrix (Σ):	
	Age	-0.0001 (0.0026)	-0.0623 [^] (0.0213)	σ_λ	0.1498 (0.0097)
	Age ²	6.24 $\cdot 10^{-6}$ (2.63 $\cdot 10^{-5}$)	6.44 $\cdot 10^{-4}$ [^] (2.11 $\cdot 10^{-4}$)	σ_r	3.1515 (0.0773)
	Female	0.0006 (0.0086)	0.2049 [^] (0.0643)	ρ	0.8391 (0.0265)
	Family			Unconditional Statistics: ¹	
	Single	omitted	omitted	Mean λ	0.2196 (0.0013)
	Married	-0.0198 (0.0115)	0.1927 [^] (0.0974)	Median λ	0.2174 (0.0017)
	Divorced	0.0396 [^] (0.0155)	-0.1754 (0.1495)	Std. Dev. λ	0.0483 (0.0019)
	Widower	0.0135 (0.0281)	-0.1320 (0.2288)	Mean r	0.0019 (0.0002)
	Other (NA)	-0.0557 (0.0968)	-0.4599 (0.7397)	Median r	7.27 $\cdot 10^{-6}$ (7.56 $\cdot 10^{-7}$)
	Education			Std. Dev. r	0.0197 (0.0015)
	Elementary	-0.0194 (0.0333)	0.1283 (0.2156)	Corr(r, λ)	0.2067 (0.0085)
	High School	omitted	omitted	Obs.	105,800
	Technical	-0.0017 (0.0189)	0.2306 (0.1341)		
	Academic	-0.0277 [^] (0.0124)	0.2177 [^] (0.0840)		
Other (NA)	-0.0029 (0.0107)	0.0128 (0.0819)			
Emigrant	0.0030 (0.0090)	0.0001 (0.0651)			
Car Attributes:	Log(Value)	0.0794 [^] (0.0177)	0.7244 [^] (0.1272)		
	Car Age	0.0053 [^] (0.0023)	-0.0411 [^] (0.0176)		
	Commercial Car	-0.0719 [^] (0.0187)	-0.0313 (0.1239)		
	Log(Engine Size)	0.1299 [^] (0.0235)	-0.3195 (0.1847)		
Driving:	License Years	-0.0015 (0.0017)	0.0157 (0.0137)		
	License Years ²	-1.83 $\cdot 10^{-5}$ (3.51 $\cdot 10^{-5}$)	-1.48 $\cdot 10^{-4}$ (2.54 $\cdot 10^{-4}$)		
	Good Driver	-0.0635 [^] (0.0112)	-0.0317 (0.0822)		
	Any Driver	0.0360 [^] (0.0105)	0.3000 [^] (0.0722)		
	Secondary Car	-0.0415 [^] (0.0141)	0.1209 (0.0875)		
	Business Use	0.0614 [^] (0.0134)	-0.3790 [^] (0.1124)		
	History Length	0.0012 (0.0052)	0.3092 [^] (0.0518)		
Claims History	0.1295 [^] (0.0154)	0.0459 (0.1670)			
Young Driver:	Young driver	0.0525 [^] (0.0253)	-0.2499 (0.2290)		
	Gender				
	Male	omitted	-		
	Female	0.0355 [^] (0.0061)	-		
	Age				
	17-19	omitted	-		
	19-21	-0.0387 [^] (0.0121)	-		
	21-24	-0.0445 [^] (0.0124)	-		
	>24	0.0114 (0.0119)	-		
Experience					
<1	omitted	-			
1-3	-0.0059 (0.0104)	-			
>3	0.0762 [^] (0.0121)	-			
Company Year:	First Year	omitted	omitted		
	Second Year	-0.0771 [^] (0.0122)	-1.4334 [^] (0.0853)		
	Third Year	-0.0857 [^] (0.0137)	-2.8459 [^] (0.1191)		
	Fourth Year	-0.1515 [^] (0.0160)	-3.8089 [^] (0.1343)		
	Fifth Year	-0.4062 [^] (0.0249)	-3.9525 [^] (0.1368)		

Standard deviations based on the draws from the posterior distribution in parentheses.

[^] Significant at the five-percent confidence level.

¹ Unconditional statistics represent implied quantities for the sample population as a whole, i.e., integrating over the distribution of covariates in the sample (as well as over the unobserved components).

Table 5: Risk aversion estimates

Specification ¹	Absolute Risk Aversion ²	Interpretation ³	Relative Risk Aversion ⁴
Back-of-the-Envelope	$1.0 \cdot 10^{-3}$	90.70	14.84
Benchmark model:			
Mean Individual	$6.7 \cdot 10^{-3}$	56.05	97.22
25th Percentile	$2.3 \cdot 10^{-6}$	99.98	0.03
Median Individual	$2.6 \cdot 10^{-5}$	99.74	0.37
75th Percentile	$2.9 \cdot 10^{-4}$	97.14	4.27
90th Percentile	$2.7 \cdot 10^{-3}$	78.34	39.02
95th Percentile	$9.9 \cdot 10^{-3}$	49.37	143.27
CARA Utility:			
Mean Individual	$3.1 \cdot 10^{-3}$	76.51	44.36
Median Individual	$3.4 \cdot 10^{-5}$	99.66	0.50
Learning Model:			
Mean Individual	$4.2 \cdot 10^{-3}$	68.86	61.40
Median Individual	$5.6 \cdot 10^{-6}$	99.95	0.08
Comparable Estimates:			
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) ⁵	$3.2 \cdot 10^{-2}$	20.96	865.75
Sydnor (2006)	$2.0 \cdot 10^{-3}$	83.29	53.95

¹ This table summarizes the results with respect to the level of risk aversion. “Back-of-the-Envelope” refers to the calculation we report in the beginning of Section 4, “Benchmark Model” refers to the results from the benchmark model (Table 4), “CARA Utility” refers to a specification of a CARA utility function, and “Learning Model” refers to a specification in which individuals do not know their risk types perfectly (see Section 4.4). The last four rows are the closest comparable results available in the literature.

² The second column presents the point estimates for the coefficient of absolute risk aversion, converted to $\$US^{-1}$ units. For the comparable estimates, 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 using the figures we report in the Mean and Median columns of risk aversion in Table 6 (multiplied by 3.52, the average exchange rate during the period, to convert to U.S. dollars).

³ To interpret the absolute risk aversion estimates (ARA), we translate them into $\{x : u(w) = \frac{1}{2}u(w + 100) + \frac{1}{2}u(w - x)\}$. That is, we report x such that an individual with the estimated ARA is indifferent about participating in a fifty-fifty lottery of gaining 100 U.S. dollars and losing x U.S. dollars. 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 back-of-the-envelope, benchmark, and learning models, and use a CARA utility function for the others.

⁴ 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. We use the average annual income (after tax) in Israel in 1995 (51,168 NIS, from Israeli census) for all our specifications, and we use average disposable income in the US in 1987 (15,437 U.S. dollars) for Gertner (1993) and Metrick (1995). For Holt and Laury (2002) and Sydnor (2006) we use a similar figure for 2002 (26,974 U.S. dollars).

⁵ 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 6: Robustness

Specification ¹	Sample	Obs.	Claim Risk (λ)			Absolute Risk Aversion (r)			Corr(r, λ)	ρ
			Mean	Median	Std. Dev.	Mean	Median	Std. Dev.		
Baseline Estimates:										
Benchmark Model	All New Customers	105,800	0.220	0.217	0.048	$1.9 \cdot 10^{-3}$	$7.3 \cdot 10^{-6}$	0.020	0.207	0.839
The vNM utility function:										
CARA Utility	All New Customers	105,800	0.219	0.217	0.048	$8.7 \cdot 10^{-4@}$	$9.8 \cdot 10^{-6@}$	$0.009@$	0.201	0.826
The claim generating process:										
Benchmark Model	No Multiple Claims	103,260	0.182	0.180	0.040	$2.0 \cdot 10^{-3}$	$2.8 \cdot 10^{-5}$	0.018	0.135	0.547
Thinner-Tail Risk Distribution	All New Customers	105,800	$0.205^{\#}$	$0.171^{\#}$	$0.155^{\#}$	$1.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-6}$	0.020	-0.076	-0.916
The distribution of risk aversion:										
Lower Bound Procedure	All New Customers	105,800	-	-	-	$3.7 \cdot 10^{-4}$	0	0.002	-	-
Incomplete information about risk:										
Benchmark Model	Experienced Drivers	82,966	0.214	0.211	0.051	$2.1 \cdot 10^{-3}$	$8.3 \cdot 10^{-6}$	0.021	0.200	0.761
Benchmark Model	Inexperienced Drivers	22,834	0.230	0.220	0.073	$3.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-7}$	0.032	0.186	0.572
Learning Model	All New Customers	105,800	0.204	0.191	0.084	$1.2 \cdot 10^{-3}$	$1.6 \cdot 10^{-6}$	0.016	0.200	0.772
Sample Selection:										
Benchmark Model	First Two Years	45,739	0.244	0.235	0.066	$3.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-5}$	0.026	0.225	0.699
Benchmark Model	Last Three Years	60,061	0.203	0.201	0.043	$1.6 \cdot 10^{-3}$	$3.4 \cdot 10^{-7}$	0.021	0.113	0.611
Benchmark Model	Referred by a Friend	26,434	0.213	0.205	0.065	$3.0 \cdot 10^{-3}$	$8.4 \cdot 10^{-7}$	0.031	0.155	0.480
Benchmark Model	Referred by Advertising	79,366	0.219	0.216	0.051	$2.1 \cdot 10^{-3}$	$7.6 \cdot 10^{-6}$	0.022	0.212	0.806
Benchmark Model	Non-Stayers	48,387	0.226	0.240	0.057	$2.3 \cdot 10^{-3}$	$7.7 \cdot 10^{-7}$	0.026	0.149	0.848
Benchmark Model	Stayers, 1st Choice	57,413	0.190	0.182	0.057	$2.9 \cdot 10^{-3}$	$2.9 \cdot 10^{-5}$	0.024	0.152	0.463
Benchmark Model	Stayers, 2nd Choice	57,413	0.208	0.200	0.065	$3.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-5}$	0.026	0.211	0.637

This table presents the key figures from various specifications and subsamples, tracing the order they are presented in Section 4.4. Full results (in the format of Table 4) from all these specifications are available in the online appendix (and at <http://www.stanford.edu/~leinav>).

¹ “Benchmark Model” refers to the benchmark specification, estimated on various subsamples (the first row replicates the estimates from Table 4). The other specifications are slightly different, and are all described in more detail in the corresponding parts of Section 4.4.

[@] The interpretation of r in the CARA model takes a slightly different quantitative meaning when applied to non-infinitesimal lotteries (such as the approximately 100 dollar stakes we analyze). This is due to the positive third derivative of the CARA utility function, compared to the benchmark model, in which we assume a small third derivative. Thus, these numbers are not fully comparable to the corresponding figures in the other specifications.

[#] The interpretation of λ in the thinner-tail distribution we estimate is slightly different from the standard Poisson rate, which is assumed in the other specifications. Thus, these numbers are not fully comparable to the corresponding figures in the other specifications.

Table 7: Representativeness

Variable		Sample ²	Population ³	Car Owners ⁴
Age ¹		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

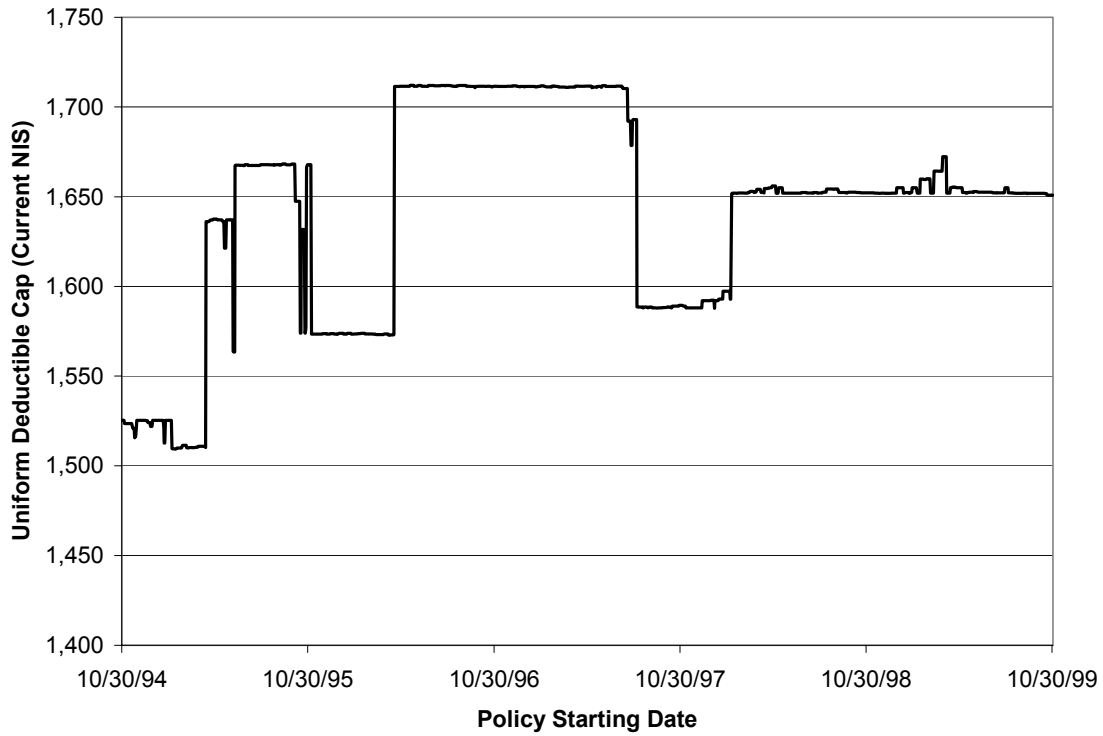
¹ For the age variable, the only continuous variable in the table, we provide both the mean and the standard deviation (in parentheses).

² The figures are derived from Table 1. The family and education variables are renormalized so they add up to one; we ignore those individuals for which we do not have family status or education level. This is particularly relevant for the education variable, which is not available for about half of the sample; it seems likely that unreported education levels are not random, but associated with lower levels of education. This may help in explaining at least some of the gap in education levels across the columns.

³ 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.

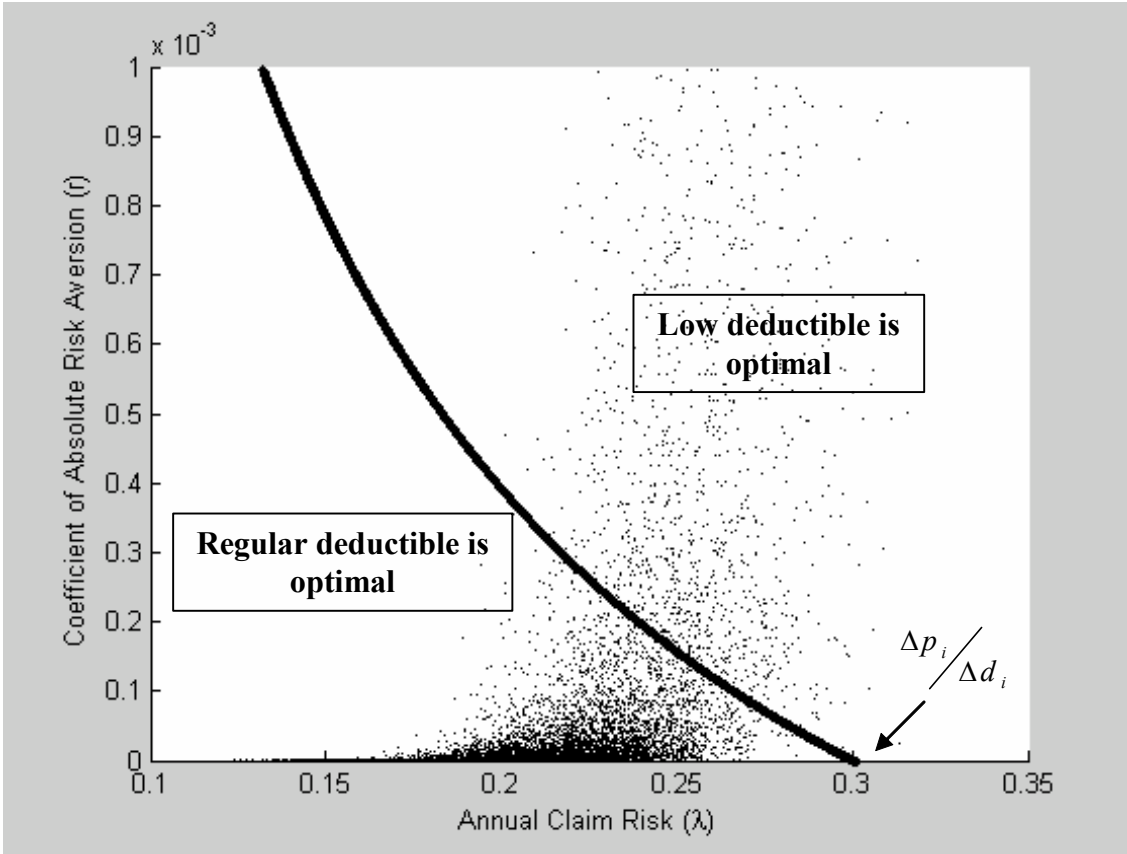
⁴ This column is based on a subsample of the population sample. The data only provide information about car ownership at the household level, not at the individual level. Thus, we define an individual as a car owner if (i) the household owns at least one car and the individual is the head of the household, 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



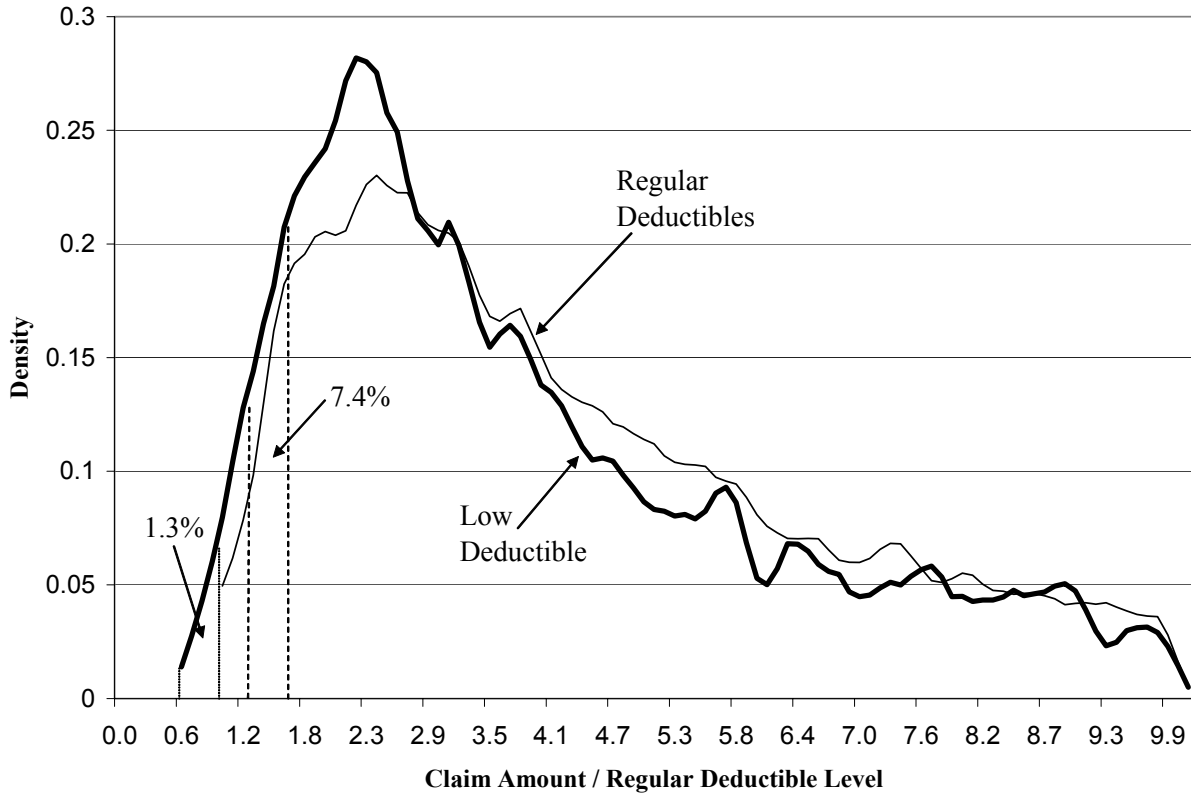
This figure presents the variation in the deductible cap over time, which is one of the main sources of pricing variation in the data. We do not observe the cap directly, but it can be calculated from the observed menus. The graph plots the maximal regular deductible offered to anyone who bought insurance from the company over a moving seven-day window. The large jumps in the graph reflect changes in the deductible cap. There are three reasons why the graph is not perfectly smooth. First, in a few holiday periods (e.g., October 1995) there are not enough sales within a seven-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 sold using earlier quotes, made according to a previous pricing regime. Finally, even within periods of constant cap, the maximal deductible varies slightly (variation of less than 0.5 percent). We do not know the source of this variation.

Figure 2: The individual’s decision – a graphical illustration



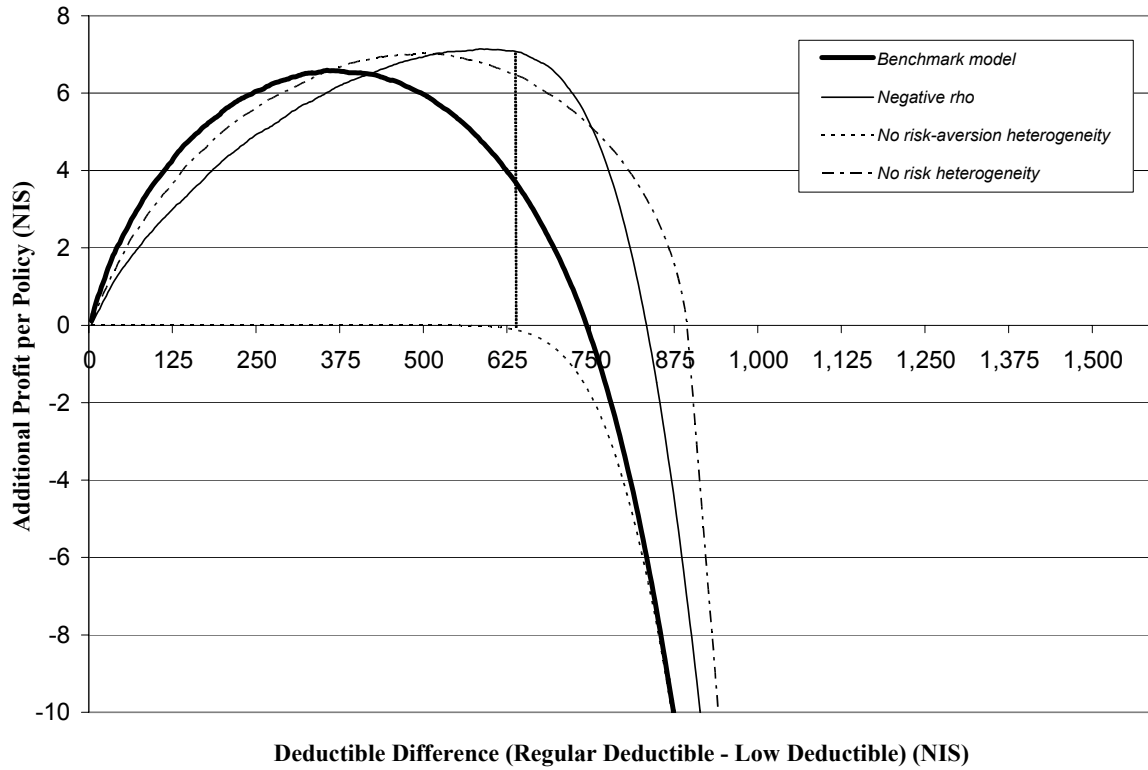
This graph illustrates the individual’s decision problem. The solid line presents the indifference set – equation (7) – applied for the menu faced by the average individual in the sample. Individuals are represented by points in the above two-dimensional space. In particular, the scattered points are 10,000 draws from the joint distribution of risk and risk aversion for the average individual (on observables) in the data, based on the point estimates of the benchmark model (Table 4). If an individual is either to the right of the line (high risk) or above the line (high risk aversion), the low deductible is optimal. Adverse selection is captured by the fact that the line is downward sloping: higher risk individuals require lower levels of risk aversion to choose the low deductible. Thus, in the absence of correlation between risk and risk aversion, higher risk individuals are more likely to choose higher levels of insurance. An individual with $\lambda_i > \frac{\Delta p_i}{\Delta d_i}$ will choose a lower deductible even if he is risk neutral, 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 that 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 \bar{d} . An increase in $\frac{\Delta p_i}{\Delta d_i}$ will shift the set up and to the right, and an increase in \bar{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,” thereby allowing us to nonparametrically identify the correlation between risk and risk aversion. With positive correlation (shown in the figure by the “right-bending” shape of the simulated draws), the marginal individuals are relatively high risk, therefore creating a stronger incentive for the insurer to raise the price of the low deductible.

Figure 3: Claim distributions



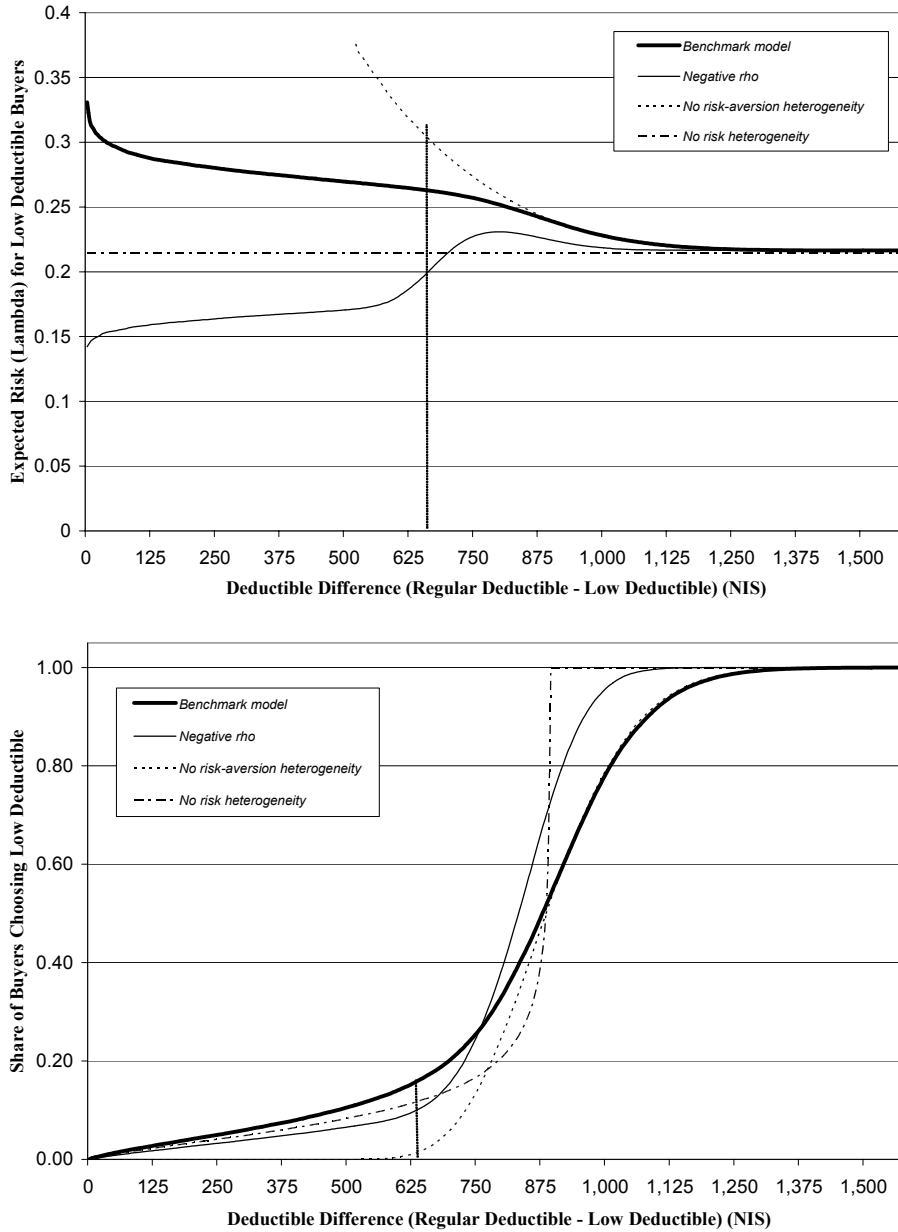
This figure plots kernel densities of the claim amounts, estimated separately, depending on the deductible choice. For ease of comparison, we normalize the claim amounts by the level of the *regular* deductible (i.e., the normalization is invariant to the deductible choice), and truncate the distribution at 10 (the truncated part, which includes a fat tail outside of the figure, accounts for about 25% of the distribution, and is roughly similar for both deductible choices). The thick line presents the distribution of the claim amounts for individuals who chose a low deductible, while the thin line does the same for those who chose a regular deductible. Clearly, both distributions are truncated from below at the deductible level. The figure shows that the distributions are fairly similar. Assuming that the claim amount distribution is the same, the area below the thicker line between 0.6 and 1 is the fraction of claims that would fall between the two deductible levels, and therefore (absent dynamic incentives) would be filed only if a low deductible was chosen. This area (between the two dotted vertical lines) amounts to 1.3 percent, implying that the potential bias arising from restricting attention to claim rate (and abstracting from the claim distribution) is quite limited. As we discuss in the text, dynamic incentives due to experience rating may increase the costs of filing a claim, shifting the region in which the deductible choice matters to the right; an upper bound to these costs is about 70 percent of the regular deductible, covering an area (between the two dashed vertical lines) that integrates to more than seven percent. Note, however, that these dynamic incentives are a very conservative upper bound; they apply to less than fifteen percent of the individuals, and do not account for the exit option, which significantly reduces these dynamic costs.

Figure 4: Counterfactuals – profits



This figure illustrates the results from the counterfactual exercise (see Section 4.6). We plot the *additional* profits from offering a low deductible as a function of the attractiveness of a low deductible, $\Delta d = d_h - d_l$, fixing its price at the observed level. The thick solid line presents the counterfactual profits as implied by the estimated benchmark model. The other three curves illustrate how the profits change in response to changes in the assumptions: when the correlation between risk and risk aversion is negative (thin solid line), when there is no heterogeneity in risk aversion (dot-dashed line), and when there is no heterogeneity in risk (dashed line). The maxima (argmax) of the four curves, respectively, are 6.59 (355), 7.14 (583), 0, and 7.04 (500). The dotted vertical line represents the observed level of Δd (638), which implies that the additional profits from the observed low deductible are 3.68 NIS per policy.

Figure 5: Counterfactuals – selection



This figure illustrates the results from the counterfactual exercise (see Section 4.6). Parallel to Figure 4, we break down the effects on profits to the share of consumer who choose a low deductible (bottom panel) and to the expected risk of this group (top panel). This is presented by the thick solid line for the estimates of the benchmark model. As in Figure 4, we also present these effects for three additional cases: when the correlation between risk and risk aversion is negative (thin solid line), when there is no heterogeneity in risk aversion (dot-dashed line), and when there is no heterogeneity in risk (dashed line). The dotted vertical lines represent the observed level of Δd (638), for which the share of low deductible is 16 percent and their expected risk is 0.264. This may be compared with the corresponding figures in Table 2A of 17.8 and 0.309, respectively. Note, however, that the figure presents estimated quantities for the average individual in the data, while Table 2A presents the average quantities in the data, so one should not expect the numbers to fit perfectly.