2012

The Nature of Risk Preferences: Evidence from Insurance Choices

Levon Barseghyan
Cornell University

Francesca Molinari
Cornell University

Joshua C. Teitelbaum
Georgetown University Law Center, jct48@law.georgetown.edu

Ted O'Donoghue
Cornell University

103 Am. Econ. Rev. 2499-2529 (2013)

The Nature of Risk Preferences: Evidence from Insurance Choices*

Levon Barseghyan†
Cornell University

Francesca Molinari
Cornell University

Ted O'Donoghue
Cornell University

Joshua C. Teitelbaum
Georgetown University

This version: November 7, 2012

Prior versions: July 26, 2012, June 30, 2011, July 21, 2010

Abstract

We use data on insurance deductible choices to estimate a structural model of risky choice that incorporates "standard" risk aversion (diminishing marginal utility for wealth) and probability distortions. We find that probability distortions—characterized by substantial overweighting of small probabilities and only mild insensitivity to probability changes—play an important role in explaining the aversion to risk manifested in deductible choices. This finding is robust to allowing for observed and unobserved heterogeneity in preferences. We demonstrate that neither Kőszegi-Rabin loss aversion alone nor Gul disappointment aversion alone can explain our estimated probability distortions, signifying a key role for probability weighting.

(*JEL* D01, D03, D12, D81, G22)

*For helpful comments, we thank seminar and conference participants at Berkeley, Collegio Carlo Alberto, Georgetown University, Harvard University, Heidelberg University, Michigan State University, Princeton University, University of Arizona, UCLA, University of Pennsylvania, Univesity of Western Ontario, the 2010 Behavioral Economics Annual Meeting, the Workshops on Behavioral and Institutional Research and Financial Regulation, FUR XIV, the 85th Annual Conference of the Western Economic Association International, the 2011 American Economic Association Annual Meeting, the 21st Annual Meeting of the American Law and Economics Association, CORIPE, and the 2011 CESifo Area Conference on Behavioural Economics. We also thank Darcy Steeg Morris for excellent research assistance. We acknowledge financial support from National Science Foundation grant SES-1031136. In addition, Barseghyan acknowledges financial support from the Institute for Social Sciences at Cornell University, O'Donoghue acknowledges financial support from CESifo, and Molinari acknowledges financial support from NSF grants SES-0617482 and SES-0922330.

†Department of Economics, Cornell University, 456 Uris Hall, Ithaca, NY 14853 (lb247@cornell.edu).

1 Introduction

Households are averse to risk—they require a premium to invest in equity and they purchase insurance at actuarially unfair rates. The standard expected utility model attributes risk aversion to a concave utility function defined over final wealth states (diminishing marginal utility for wealth). Indeed, many empirical studies of risk preferences assume expected utility and estimate such "standard" risk aversion (e.g., Cohen and Einav 2007).

A considerable body of research, however, suggests that in addition to (or perhaps instead of) standard risk aversion, households' aversion to risk may be attributable to other, "non-standard" features of risk preferences. A large strand of the literature focuses on probability weighting and loss aversion, two features that originate with prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992). Alternatively, Gul (1991) and others propose models that feature various forms of disappointment aversion. More recently, Kőszegi and Rabin (2006, 2007) develop a model of reference-dependent preferences that features a form of "rational expectations" loss aversion (which we label KR loss aversion).

In this paper, we use data on households' deductible choices in auto and home insurance to estimate a structural model of risky choice that incorporates standard risk aversion and these non-standard features, and we investigate which combinations of features best explain our data. We show that, in our domain, probability weighting, KR loss aversion, and Gul disappointment aversion all imply an effective *distortion* of probabilities relative to the expected utility model. Hence, we focus on estimating a model that features standard risk aversion and "generic" probability distortions. We then investigate what we can learn from our estimates about the possible sources of probability distortions. We find that probability distortions—in the form of substantial overweighting of claim probabilities—play a key role in explaining households' deductible choices. We then demonstrate that neither KR loss aversion alone nor Gul disappointment aversion alone can explain our estimated probability distortions, signifying a crucial role for probability weighting.

In Section 2, we provide an overview of our data. The source of the data is a large U.S. property and casualty insurance company that offers multiple lines of insurance, including auto and home coverage. The full dataset comprises yearly information on more than 400,000 households who held auto or home policies between 1998 and 2006. For reasons we explain, we restrict attention in our main analysis to a core sample of 4170 households who hold both auto and home policies and who first purchased their auto and home policies from the company in the same year, in either 2005 or 2006. For each household, we observe the household's deductible choices in three lines of coverage—auto collision, auto comprehensive, and home all perils. We also observe the coverage-specific menus of premium-deductible combinations from which each household's choices were made. In addition, we observe each household's claims history for each coverage, as well as a rich set of demographic information. We utilize the data on claim realizations and demographics to assign each household a predicted claim probability for each coverage.

In Section 3, we develop our theoretical framework. We begin with an expected utility model of deductible choice, which incorporates standard risk aversion. We then generalize the model to allow for probability distortions—specifically, we permit a household with claim probability μ to act as if its claim probability were $\Omega(\mu)$. In our baseline analysis, we take a semi-nonparametric approach, and do not impose a parametric form on the probability distortion function $\Omega(\mu)$. For the utility for wealth function, we use a second-order Taylor expansion, which allows us to measure standard risk aversion by the coefficient of absolute risk aversion r. Finally, to account for observationally equivalent households choosing different deductibles, and for individual households making "inconsistent" choices across coverages (Barseghyan et al. 2011; Einav et al. 2012), we assume random utility with additively separable choice noise (McFadden 1974, 1981).

In Section 3.3, we demonstrate that a key feature of our data—namely, that the choice set for each coverage includes more than two deductible options—enables us to separately identify standard risk aversion r and the probability distortion $\Omega(\mu)$. To illustrate the basic intuition, consider a household with claim probability μ, and suppose we observe the household's maximum willingness to pay (WTP) to reduce its deductible from \$1000 to \$500. If that were all we observed, we could not separately identify r and $\Omega(\mu)$, because multiple combinations can explain this WTP. However, because standard risk aversion and probability distortions generate aversion to risk in different ways, each of these combinations implies a different WTP to further reduce the deductible from \$500 to \$250. Therefore, if we also observe this WTP, we can pin down r and $\Omega(\mu)$.

In Section 4, we report the results of our baseline analysis in which we assume homogenous preferences—i.e., we assume that each household has the same standard risk aversion r and the same probability distortion function $\Omega(\mu)$. We take three approaches based on the method of sieves (Chen 2007) to estimating $\Omega(\mu)$, each of which yields the same main message: large probability distortions, characterized by substantial overweighting of claim probabilities and only mild insensitivity to probability changes, in the range of our data. Under our primary approach, for example, our estimates imply $\Omega(0.02) = 0.08$, $\Omega(0.05) = 0.11$, and $\Omega(0.10) = 0.16$. In Section 4.2, we demonstrate the statistical and economic significance of our estimated $\Omega(\mu)$.

In Section 4.3, we discuss what we learn from our baseline estimates about the possible sources of probability distortions. We briefly describe models of probability weighting, KR loss aversion, and Gul disappointment aversion, and derive the probability distortion function implied by each model.[1] We demonstrate that models of KR loss aversion and of Gul disappointment aversion imply probability distortions that are inconsistent with our estimated $\Omega(\mu)$. We therefore conclude that we can "reject" the hypothesis that KR loss aversion alone or Gul disappointment aversion alone is the source of our estimated probability distortions, and that instead our results point to probability weighting. In addition, we highlight that our estimated $\Omega(\mu)$ bears a close resemblance to the probability weighting function originally posited by Kahneman and Tversky (1979).

In Section 5, we expand the model to permit heterogeneous preferences—i.e., we allow each household to have a different combination of standard risk aversion and probability distortions. We take three approaches, permitting first only observed heterogeneity, then only unobserved heterogeneity, and then both observed and unobserved heterogeneity.[2] We find that our main message is robust to allowing for heterogeneity in preferences. While our estimates indicate substantial heterogeneity, under each approach the average probability distortion function is remarkably similar to our baseline estimated $\Omega(\mu)$.

In Section 6, we investigate the sensitivity of our estimates to other modeling assumptions. Most notably, we extend the model to account for unobserved heterogeneity in claim probabilities, we consider the case of constant relative risk aversion (CRRA) utility, and we address the issue of moral hazard. All in all, we find that our main message is quite robust.

We conclude in Section 7 by discussing certain implications and limitations of our study. Among other things, we discuss the relevance of the fact that, in our data, probability weighting is indistinguishable from systematic risk misperceptions. Hence, the probability distortions we estimate may reflect either probability weighting or risk misperceptions.

Numerous previous studies estimate risk preferences from observed choices, relying in most cases on survey and experimental data and in some cases on economic field data. Most studies that rely on field data—including two that use data on insurance deductible choices (Cohen and Einav 2007; Sydnor 2010)—estimate expected utility models, which permit only standard risk aversion. Only a handful of studies use field data to estimate models that feature probability weighting. Cicchetti and Dubin (1994), who use data on telephone wire insurance choices to estimate a rank-dependent expected utility model, and Jullien and Salanié (2000), who use data on bets on U.K. horse races to estimate a rank-dependent expected utility model and a prospect theory model, find little evidence of probability weighting. Kliger and Levy (2009) use data on call options on the S&P 500 index to estimate a prospect the-

[1] Detailed descriptions of these models appear in the Appendix.

[2] A number of recent papers have studied the role of unobserved heterogeneity in risk preferences (e.g., Cohen and Einav 2007; Chiappori et al. 2009; Andrikogiannopoulou 2010).

ory model, and find that probability weighting is manifested by their data. Snowberg and Wolfers (2010) use data on bets on U.S. horse races to test the fit of two models—a model with standard risk aversion alone and a model with probability weighting alone—and find that the latter model better fits their data. Andrikogiannopoulou (2010) uses data on bets in an online sportsbook to estimate a prospect theory model, and finds that the average bettor exhibits moderate probability weighting. Lastly, Chiappori et al. (2012) and Gandhi and Serrano-Padial (2012) use data on bets on U.S. horse races to estimate the distribution of risk preferences among bettors, and find evidence consistent with probability weighting.[3]

Each of the foregoing studies, however, either (i) uses market-level data, which necessitates taking a representative agent approach (Jullien and Salanié 2000; Kliger and Levy 2009; Snowberg and Wolfers 2010) or assuming that different populations of agents have the same distribution of risk preferences (Chiappori et al. 2012; Gandhi and Serrano-Padial 2012), (ii) estimates a model that does not simultaneously feature standard risk aversion and probability weighting (Kliger and Levy 2009; Snowberg and Wolfers 2010; Andrikogiannopoulou 2010),[4] or (iii) takes a parametric approach to probability weighting, specifying one of the common functions from the literature (Cicchetti and Dubin 1994; Jullien and Salanié 2000; Kliger and Levy 2009; Andrikogiannopoulou 2010). An important contribution of our study is that we use household-level field data on insurance choices to jointly estimate standard risk aversion and non-standard probability distortions without imposing a parametric form on the latter.[5] Our approach in this regard yields two important results. First, by imposing no parametric form on $\Omega(\mu)$, we estimate a function that is inconsistent with the probability weighting functions that are commonly used in the literature (e.g., Tversky and Kahneman 1992; Lattimore et al. 1992; Prelec 1998). Second, by jointly estimating r and $\Omega(\mu)$, we can empirically assess their relative impact on choices, and we find that probability distortions generally have a larger economic impact than standard risk aversion.

[3]Bruhin et al. (2010) is another recent study that echoes our conlcusion that probability weighting is important. They use experimental data on subjects' choices over binary money lotteries to estimate a mixture model of cumulative prospect theory. They find that approximately 20 percent of subjects can essentially be characterized as expected value maximizers, while approximately 80 percent exhibit significant probability weighting.

[4]As noted above, Snowberg and Wolfers (2010) estimate two separate models—one with standard risk aversion alone and one with probability weighting alone. Kliger and Levy (2009) and Andrikogiannopoulou (2010) estimate cumulative prospect theory models, which feature a value function defined over gains and losses in lieu of the standard utility function defined over final wealth states. Both studies, incidentally, find evidence of "status quo" loss aversion.

[5]This is the case for our baseline analysis in Section 4. In our analysis in Sections 5 and 6, we take a parametric approach that is guided by the results of our baseline analysis.

2 Data Description

2.1 Overview and Core Sample

We acquired the data from a large U.S. property and casualty insurance company. The company offers multiple lines of insurance, including auto, home, and umbrella policies. The full dataset comprises yearly information on more than 400,000 households who held auto or home policies between 1998 and 2006.[6] For each household, the data contain all the information in the company's records regarding the households and their policies (except for identifying information). The data also record the number of claims that each household filed with the company under each of its policies during the period of observation.

We restrict attention to households' deductible choices in three lines of coverage: (i) auto collision coverage; (ii) auto comprehensive coverage; and (iii) home all perils coverage.[7] In addition, we consider only the initial deductible choices of each household. This is meant to increase confidence that we are working with active choices; one might be concerned that some households renew their policies without actively reassessing their deductible choices. Finally, we restrict attention to households who hold both auto and home policies and who first purchased their auto and home policies from the company in the same year, in either 2005 or 2006. The latter restriction is meant to avoid temporal issues, such as changes in household characteristics and in the economic environment. In the end, we are left with a core sample of 4170 households. Table 1 provides descriptive statistics for a subset of variables, specifically those we use later to estimate the households' utility parameters.

TABLE 1

2.2 Deductibles and Premiums

For each household in the core sample, we observe the household's deductible choices for auto collision, auto comprehensive, and home, as well as the premiums paid by the household for each type of coverage. In addition, the data contain the exact menus of premium-deductible combinations that were available to each household at the time it made its deductible choices. Tables 2 and 3 summarize the deductible choices and the premium menus, respectively, of

[6] The dataset used in this paper is not the same dataset used in Barseghyan et al. (2011). This dataset includes households that purchase insurance from a single insurance company (through multiple insurance agents), whereas that dataset includes households that purchase insurance through a single insurance agent (from multiple insurance companies).

[7] A brief description of each type of coverage appears in the Appendix. For simplicity, we often refer to home all perils merely as home.

the households in the core sample.[8]

TABLES 2 & 3

Because it is important to understand the sources of variation in premiums, we briefly describe the plan the company uses to rate a policy in each line of coverage. We emphasize that the company's rating plan is subject to state regulation and oversight. In particular, the regulations require that the company receive prior approval of its rating plan by the state insurance commissioner, and they prohibit the company and its agents from charging rates that depart from the plan. Under the plan, the company determines a base price $\bar{p}$ for each household according to a coverage-specific rating function, which takes into account the household's coverage-relevant characteristics and any applicable discounts. Using the base price, the company then generates a household-specific menu $\{(p_d, d) : d \in \mathcal{D}\}$, which associates a premium p_d with each deductible d in the coverage-specific set of deductible options $\mathcal{D}$, according to a coverage-specific multiplication rule, $p_d = (g(d) \cdot \bar{p}) + c$, where $g(\cdot) > 0$ and $c > 0$. The multiplicative factors $\{g(d) : d \in \mathcal{D}\}$ are known in the industry as the deductible factors and c is a small markup known as the expense fee. The deductible factors and the expense fee are coverage specific but household invariant.

2.3 Claim Probabilities

For purposes of our analysis, we need to estimate for each household the likelihood of experiencing a claim for each coverage. We begin by estimating how claim rates depend on observables. In an effort to obtain the most precise estimates, we use the full dataset: 1,348,020 household-year records for auto and 1,265,229 household-year records for home. For each household-year record, the data record the number of claims filed by the household in that year. We assume that household i's claims under coverage j in year t follow a Poisson distribution with arrival rate λ_{ijt}. In addition, we assume that deductible choices do not influence claim rates, i.e., households do not suffer from moral hazard.[9] We treat the claim rates as latent random variables and assume that

$$\ln \lambda_{ijt} = X'_{ijt}\beta_j + \epsilon_{ij},$$

[8] Tables A.1 through A.3 in the Appendix summarize the premium menus conditional on households' actual deductible choices.

[9] We revisit this assumption in Section 6.4.

where X_{ijt} is a vector of observables,[10] ϵ_{ij} is an unobserved iid error term, and $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance ϕ_j. We perform standard Poisson panel regressions with random effects to obtain maximum likelihood estimates of β_j and ϕ_j for each coverage j. By allowing for unobserved heterogeneity, the Poisson random effects model accounts for overdispersion, including due to excess zeros, in a similar way as the (pooled) negative binomial model (see, e.g., Wooldridge 2002, ch. 19).[11] The results of the claim rate regressions are reported in Tables A.4 and A.5 in the Appendix.

Next, we use the results of the claim rate regressions to generate predicted claim probabilities. Specifically, for each household i, we use the regression estimates to generate a predicted claim rate $\widehat{\lambda}_{ij}$ for each coverage j, conditional on the household's ex ante characteristics X_{ij} and ex post claims experience.[12] In principle, during the policy period, a household may experience zero claims, one claim, two claims, and so forth. In the model, we assume that households disregard the possibility of more than one claim (see Section 3.1).[13] Given this assumption, we transform $\widehat{\lambda}_{ij}$ into a predicted claim probability $\widehat{\mu}_{ij}$ using[14]

$$\widehat{\mu}_{ij} = 1 - \exp(-\widehat{\lambda}_{ij}).$$

Table 4 summarizes the predicted claim probabilities for the core sample. Figure 1 plots the empirical density functions. The mean predicted claim probabilities for auto collision, auto comprehensive, and home are 0.069, 0.021, and 0.084, respectively. Auto comprehensive accounts for most of the low claim probabilities, while auto collision and home account for the bulk of the medium and high claim probabilities. Table 4 also reports pairwise correlations among the predicted claim probabilities and between the predicted claim probabilities and the premiums for coverage with a \$500 deductible. Each of the pairwise correlations is positive, though none are large. These small correlations are not surprising. Even for a fixed claim probability, there are a variety of reasons why the company would want to charge different premiums to different households—e.g., differences in the insured value of the

[10] In addition to the variables in Table 1 (which we use later to estimate the households' utility parameters), X_{ijt} includes numerous other variables (see Tables A.4 and A.5 in the Appendix).

[11] An alternative approach would be a zero-inflated model. However, Vuong (1989) and likelihood ratio tests select the negative binomial model over the zero-inflated model, suggesting that adjustment for excess zeros is not necessary once we allow for unobserved heterogeneity.

[12] More specifically, $\widehat{\lambda}_{ij} = \exp(X'_{ij}\widehat{\beta}_j)E(\exp(\epsilon_{ij})|Y_{ij})$, where Y_{ij} records household i's claims experience under coverage j after purchasing the policy and $E(\exp(\epsilon_{ij})|Y_{ij})$ is calculated assuming $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance $\widehat{\phi}_j$.

[13] Because claim rates are small (85 percent of the predicted claim rates in the core sample are less than 0.1, and 99 percent are less than 0.2), the likelihood of two or more claims is very small.

[14] The Poisson probability mass function is $f(x, \lambda) = \exp(-\lambda)\lambda^x/x!$ for $x = 0, 1, 2, ...$ and $\lambda \geq 0$. Thus, if the number of claims x follows a Poisson distribution with arrival rate λ, then the probability of experiencing at least one claim is $1 - \exp(-\lambda)$.

auto or home, differences in the relevant repair and rebuilding costs, and volume discounts. Moreover, our predicted claim probabilities take into account ex post claims experience (i.e., claims that occur after the household purchases the policy), and this information is not available to the company when it rates the policy.

TABLE 4 & FIGURE 1

3 Model, Estimation, and Identification

3.1 Model

We assume that a household treats its three deductible choices as independent decisions. This assumption is motivated in part by computational considerations,[15] but also by the literature on "narrow bracketing" (e.g., Read et al. 1999), which suggests that when people make multiple choices, they frequently do not assess the consequences in an integrated way, but rather tend to make each choice in isolation. Thus, we develop a model for how a household chooses the deductible for a single type of insurance coverage. To simplify notation, we suppress the subscripts for household and coverage (although we remind the reader that premiums and claim probabilities are household and coverage specific).

The household faces a menu of premium-deductible pairs $\{(p_d, d) : d \in \mathcal{D}\}$, where p_d is the premium associated with deductible d, and $\mathcal{D}$ is the set of deductible options. We assume that the household disregards the possibility of experiencing more than one claim during the policy period, and that the probability of experiencing one claim is μ. In addition, we assume that the household believes that its choice of deductible does not influence its claim probability, and that every claim exceeds the highest available deductible.[16] Under the foregoing assumptions, the choice of deductible involves a choice among *deductible lotteries* of the form

$$L_d \equiv (-p_d, 1-\mu; -p_d - d, \mu) \, .$$

Under expected utility theory, a household's preferences over deductible lotteries are influenced only by standard risk aversion. Given initial wealth w, the expected utility of

[15] If instead we were to assume that a household treats its deductible choices as a joint decision, then the household would face 120 options and the utility function would have several hundred terms.

[16] We make the latter assumption more plausible by excluding the $2500 and $5000 deductible options from the home menu. Only 1.6 percent of households in the core sample chose a home deductible of $2500 or $5000. We assign these households a home deductible of $1000. In this respect, we follow Cohen and Einav (2007), who also exclude the two highest deductible options (chosen by 1 percent of the policyholders in their sample) and assign the third highest deductible to policyholders who chose the two highest options.

deductible lottery L_d is given by

$$EU(L_d) = (1-\mu)\,u\,(w-p_d) + \mu u\,(w-p_d-d)\,,$$

where $u(w)$ represents standard utility defined over final wealth states. Standard risk aversion is captured by the concavity of $u(w)$.

Over the years, economists and other social scientists have proposed alternative models that feature additional sources of aversion to risk. Kahneman and Tversky (1979) and Tversky and Kahneman (1992) offer prospect theory, which features probability weighting and loss aversion. Gul (1991) proposes a model of disappointment aversion. More recently, Kőszegi and Rabin (2006, 2007) develop a model of reference-dependent utility that features loss aversion with an endogenous reference point, which we label KR loss aversion. In the Appendix, we show that, in our setting, probability weighting, KR loss aversion, and Gul disappointment aversion all imply an effective *distortion* of probabilities relative to the expected utility model.[17] Specifically, each implies that there exists a *probability distortion function* $\Omega\,(\mu)$ such that the utility of deductible lottery L_d may be written as

$$U(L_d) = (1-\Omega\,(\mu))\,u\,(w-p_d) + \Omega\,(\mu)\,u\,(w-p_d-d)\,. \tag{1}$$

In the estimation, we do not take a stand on the underlying source of probability distortions. Their separate identification would require parametric assumptions which we are unwilling to make. Rather, we focus on estimating "generic" probability distortions—i.e., we estimate the function $\Omega\,(\mu)$. We then discuss what we learn from our estimated $\Omega\,(\mu)$ about the possible sources of probability distortions (see Section 4.3).

In our analysis, we estimate both the utility function $u(w)$ and the probability distortion function $\Omega\,(\mu)$. For $u(w)$, we generally follow Cohen and Einav (2007) and Barseghyan et al. (2011) and consider a second-order Taylor expansion. Also, because $u(w)$ is unique only up to an affine transformation, we normalize the scale of utility by dividing $u'(w)$. This yields

$$\frac{u(w+\Delta)}{u'(w)} - \frac{u(w)}{u'(w)} = \Delta - \frac{r}{2}\Delta^2,$$

where $r \equiv -u''(w)/u'(w)$ is the coefficient of absolute risk aversion. Because the term $u(w)/u'(w)$ enters as an additive constant, it does not affect utility comparisons; hence, we

[17] We do not consider the original, "status quo" loss aversion proposed by Kahneman and Tversky (1979) because it cannot explain aversion to risk in the context of insurance deductible choices, where all outcomes are losses relative to initial wealth. Instead, we consider KR loss aversion, which can explain aversion to risk in this context, because gains and losses are defined relative to expectations about outcomes. For details, see Section 4.3 and the Appendix.

drop it. With this specification, equation (1) becomes

$$U(L_d) = -\left[p_d + \Omega(\mu)d\right] - \frac{r}{2}\left[\left(1-\Omega(\mu)\right)(p_d)^2 + \Omega(\mu)\left(p_d+d\right)^2\right]. \tag{2}$$

The first term reflects the expected value of deductible lottery L_d with respect to the distorted claim probability $\Omega(\mu)$. The second term reflects disutility from bearing risk—it is the expected value of the squared losses, scaled by standard risk aversion r.[18]

Our goal is to estimate both standard risk aversion r and and the probability distortion $\Omega(\mu)$. In our baseline analysis in Section 4, we take a semi-nonparametric approach and estimate $\Omega(\mu)$ via sieve methods (Chen 2007). In Sections 5 and 6, we take a parametric approach that is guided by the results of our baseline analysis.

3.2 Estimation

In the estimation, we must account for observationally equivalent households choosing different deductibles, and for individual households making "inconsistent" choices across coverages (Barseghyan et al. 2011; Einav et al. 2012). We follow McFadden (1974, 1981) and assume random utility with additively separable choice noise. Specifically, we assume that the utility from deductible $d \in \mathcal{D}$ is given by

$$\mathcal{U}(d) \equiv U(L_d) + \varepsilon_d, \tag{3}$$

where ε_d is an iid random variable that represents error in evaluating utility. We assume that ε_d follows a type 1 extreme value distribution with scale parameter σ.[19] A household chooses deductible d when $\mathcal{U}(d) > \mathcal{U}(d')$ for all $d' \neq d$, and thus the probability that a household chooses deductible d is

$$\Pr(d) \equiv \Pr\left(\varepsilon_{d'} - \varepsilon_d < U(L_d) - U(L_{d'})\ \forall\ d' \neq d\right) = \frac{\exp\left(U(L_d)/\sigma\right)}{\sum_{d'\in\mathcal{D}}\exp\left(U(L_{d'})/\sigma\right)}.$$

We use these choice probabilities to construct the likelihood function in the estimation.

In our main analysis, we estimate equation (3) assuming that utility is specified by equation (2) and that $\mu_{ij} = \widehat{\mu}_{ij}$ (i.e., household i's predicted claim probability $\widehat{\mu}_{ij}$ corresponds to it subjective claim probability μ_{ij}). We use combined data for all three coverages. Each

[18] Note that this specification differs slightly from Cohen and Einav (2007) and Barseghyan et al. (2011), who use $U(L_d) = -\left[p_d + \lambda d\right] - \frac{r}{2}\left[\lambda d^2\right]$ (where λ is the Poisson arrival rate). The difference derives from the fact that those papers additionally take the limit as the policy period becomes arbitrarily small.

[19] The scale parameter σ is a monotone transformation of the variance of ε_d, and thus a larger σ means larger variance. Our estimation procedure permits σ to vary across coverages.

observation comprises, for a household i and a coverage j, a deductible choice d^*_{ij}, a vector of household characteristics Z_i, a predicted claim probability $\widehat{\mu}_{ij}$, and a menu of premium-deductible combinations $\{(p_{d_{ij}}, d_{ij}) : d_{ij} \in \mathcal{D}_j\}$. To be estimated are:

r_i	–	coefficient of absolute risk aversion ($r_i = 0$ means no standard risk aversion);
$\Omega_i(\mu)$	–	probability distortion function ($\Omega_i(\mu) = \mu$ means no probability distortions); and
σ_j	–	scale of choice noise for coverage j ($\sigma_j = 0$ means no choice noise).

Note that because the scale of utility is pinned down in equation (2), we can identify σ_L, σ_M, and σ_H separately for auto collision, auto comprehensive, and home, respectively.

3.3 Identification

3.3.1 Identifying r and $\Omega(\mu)$

The random utility model in equation (3) comprises the sum of a utility function $U(L_d)$ and an error term ε_d. Using the results of Matzkin (1991), normalizations that fix scale and location, plus regularity conditions that are satisfied in our model, allow us to identify nonparametrically the utility function $U(L_d)$ within the class of monotone and concave utility functions. As we explain below, identification of $U(L_d)$ allows us to identify r and $\Omega(\mu)$.

Take any three deductible options $a, b, c \in \mathcal{D}$, with $a > b > c$, and consider a household with premium p_a for deductible a and claim probability μ. The household's r and $\Omega(\mu)$ determine the premium $\widetilde{p}_b$ that makes the household indifferent between deductibles a and b, as well as the premium $\widetilde{p}_c$ that makes the household indifferent between deductibles a and c. Notice that $\widetilde{p}_b - p_a$ reflects the household's maximum willingness to pay (WTP) to reduce its deductible from a to b, and $\widetilde{p}_c - \widetilde{p}_b$ reflects the household's additional WTP to reduce its deductible from b to c. In the Appendix, we prove the following properties of $\widetilde{p}_b$ and $\widetilde{p}_c$ when $U(L_d)$ is specified by equation (2).

Property 1. *Both $\widetilde{p}_b$ and $\widetilde{p}_c$ are strictly increasing in r and $\Omega(\mu)$.*

Property 2. *For any fixed $\Omega(\mu)$, $r = 0$ implies $\frac{\widetilde{p}_b - p_a}{\widetilde{p}_c - \widetilde{p}_b} = \frac{a-b}{b-c}$, and the ratio $\frac{\widetilde{p}_b - p_a}{\widetilde{p}_c - \widetilde{p}_b}$ is strictly increasing in r.*

Property 3. *If $\widetilde{p}_b - p_a$ is the same for $(r, \Omega(\mu))$ and $(r', \Omega(\mu)')$ with $r < r'$ (and thus $\Omega(\mu) > \Omega(\mu)'$), then $\widetilde{p}_c - \widetilde{p}_b$ is greater for $(r, \Omega(\mu))$ than for $(r', \Omega(\mu)')$.*

Property 1 is straightforward: A household's WTP to reduce its deductible (and thereby reduce its exposure to risk) will be greater if either its standard risk aversion is larger or its (distorted) claim probability is larger.

Property 2 is an implication of standard risk aversion. For any fixed $\Omega(\mu)$, a risk neutral household is willing to pay, for instance, exactly twice as much to reduce its deductible from \$1000 to \$500 as it is willing to pay to reduce its deductible from \$500 to \$250. In contrast, a risk averse household is willing to pay more than twice as much, and the larger is the household's standard risk aversion the greater is this ratio.

Property 3 is the key property for identification. To illustrate the underlying intuition, consider a household with claim probability $\mu = 0.05$ who faces a premium $p_a = \$200$ for deductible $a = \$1000$. Suppose that $\widetilde{p}_b - p_a = \$50$—i.e., the household's WTP to reduce its deductible from $a = \$1000$ to $b = \$500$ is fifty dollars. Property 1 implies that multiple combinations of r and $\Omega(0.05)$ are consistent with this WTP, and that in each combination a larger r implies a smaller $\Omega(0.05)$. For example, both (i) $r = 0$ and $\Omega(0.05) = 0.10$ and (ii) $r' = 0.00222$ and $\Omega(0.05)' = 0.05$ are consistent with $\widetilde{p}_b - p_a = \$50$. Property 3, however, states that these different combinations of r and $\Omega(0.05)$ have different implications for the household's WTP to reduce its deductible from $b = \$500$ to $c = \$250$. For instance, $r = 0$ and $\Omega(0.05) = 0.10$ would imply $\widetilde{p}_c - \widetilde{p}_b = \$25$, whereas $r' = 0.00222$ and $\Omega(0.05)' = 0.05$ would imply $\widetilde{p}_c - \widetilde{p}_b = \$18.61$. More generally—given the household's WTP to reduce its deductible from \$1000 to \$500—the smaller is the household's WTP to reduce it deductible from \$500 to \$250, the larger must be its r and the smaller must be its $\Omega(0.05)$.[20]

Property 3 reveals that our identification strategy relies on a key feature of our data—namely, that there are more than two deductible options. Given this feature, we can separately identify r and $\Omega(\mu)$ by observing how deductible choices react to exogenous changes in premiums for a fixed claim probability.[21] We then can identify the shape of $\Omega(\mu)$ by observing how deductible choices react to exogenous changes in claim probabilities. In other words, given three or more deductible options, it is exogenous variation in premiums for a fixed μ that allows us to pin down r and $\Omega(\mu)$, and it is exogenous variation in claim probabilities that allows us to map out $\Omega(\mu)$ for all μ in the range of our data.

3.3.2 Exogenous Variation in Premiums and Claim Probabilities

Within each coverage, there is substantial variation in premiums and claim probabilities. A key identifying assumption is that there is variation in premiums and claim probabilities that is exogenous to the households' risk preferences. In our estimation, we assume that a household's utility parameters—r and $\Omega(\mu)$—depend on a vector of observables Z that is a strict subset of the variables that determine premiums and claim probabilities. Many of the

[20] Property 3 reads like a single crossing property. This is noteworthy in light of recent work that relies on a single crossing condition to estimate risk preferences (e.g., Chiappori et al. 2012).

[21] Note that if the data included only two deductible options, as in Cohen and Einav (2007), we could not separately identify r and $\Omega(\mu)$ without making strong functional form assumptions about $\Omega(\mu)$.

variables outside Z that determine premiums and claim probabilities, such as protection class and territory code,[22] are arguably exogenous to the households' risk preferences. In addition, there are other variables outside Z that determine premiums but not claim probabilities, including numerous discount programs, which also are arguably exogenous to the households' risk preferences.

Given our choice of Z, there is substantial variation in premiums and claim probabilities that is not explained by Z. In particular, regressions of premiums and predicted claim probabilities on Z yield low coefficients of determination (R^2). In the case of auto collision coverage, for example, regressions of premiums (for coverage with a \$500 deductible) on Z and predicted claim probabilities on Z yield R^2 of 0.16 and 0.34, respectively.[23]

In addition to the substantial variation in premiums and claim probabilities within a coverage, there also is substantial variation in premiums and claim probabilities across coverages. A key feature of the data is that for each household we observe deductible choices for three coverages, and (even for a fixed Z) there is substantial variation in premiums and claim probabilities across the three coverages. Thus, even if the within-coverage variation in premiums and claim probabilities were insufficient in practice, we still might be able to estimate the model using across-coverage variation.

4 Analysis with Homogenous Preferences

We begin our analysis by assuming homogeneous preferences—i.e., r and $\Omega(\mu)$ are the same for all households. This allows us to take a semi-nonparametric approach to estimating the model without facing a curse of dimensionality. As a point of reference, we note that if we do not allow for probability distortions—i.e., we restrict $\Omega(\mu) = \mu$—the estimate for r is 0.0129 (standard error: 0.0004).

4.1 Estimates

We take three sieve approaches to estimating $\Omega(\mu)$. In Model 1a, we estimate a Chebyshev polynomial expansion of $\ln \Omega(\mu)$, which naturally constrains $\Omega(\mu) > 0$. We consider expansions up to the 20th degree, and select a quadratic on the basis of the Bayesian information criterion (BIC). In Model 1b, we estimate a Chebyshev polynomial expansion of

[22] Protection class gauges the effectiveness of local fire protection and building codes. Territory codes are based on actuarial risk factors, such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services.

[23] They are even lower for auto comprehensive and home. In the case of auto comprehensive the coefficients of determination are 0.07 and 0.31, and in the case of home they are 0.04 and 0.15.

$\Omega(\mu)$, which nests the case $\Omega(\mu) = \mu$.[24] As before, we consider expansions up to the 20th degree. Here, the BIC selects a cubic. However, because the BIC for the quadratic and cubic are essentially the same, we report results for the quadratic to facilitate direct comparisons with Model 1a. In Model 1c, we estimate $\Omega(\mu)$ using an 11-point cubic spline on the interval $(0, 0.20)$, wherein lie 99.4 percent of the predicted claim probabilities in the core sample.

Table 5 reports our results. The estimates for $\Omega(\mu)$ indicate large probability distortions. To illustrate, Figure 2 depicts the estimated $\Omega(\mu)$ for Models 1a, 1b, and 1c, along with the 95 percent pointwise bootstrap confidence bands for Model 1c.[25] In each model, there is substantial overweighting of claim probabilities. In Model 1a, for example, $\widehat{\Omega}(0.020) = 0.083$, $\widehat{\Omega}(0.050) = 0.111$, $\widehat{\Omega}(0.075) = 0.135$, and $\widehat{\Omega}(0.100) = 0.156$. In addition, there is only mild insensitivity to probability changes. For instance, in Model 1a, $[\widehat{\Omega}(0.050) - \widehat{\Omega}(0.020)]/[0.050 - 0.020] = 0.933$, $[\widehat{\Omega}(0.075) - \widehat{\Omega}(0.050)]/[0.075 - 0.050] = 0.960$, and $[\widehat{\Omega}(0.100) - \widehat{\Omega}(0.075)]/[0.100 - 0.075] = 0.840$. Moreover, all three models imply nearly identical distortions of claim probabilities between zero and 14 percent (wherein lie 96.7 percent of the predicted claim probabilities in the core sample), and even for claim probabilities greater than 14 percent the three models are statistically indistinguishable (Models 1a and 1b lie within the 95 percent confidence bands for Model 1c). Given this overweighting, the estimates for r are smaller than without probability distortions. Specifically, $\widehat{r}$ is 0.00064, 0.00063, and 0.00049 in Models 1a, 1b, and 1c, respectively. Lastly, we note the estimates for the scale of choice noise: $\widehat{\sigma}_L = 26.3$, $\widehat{\sigma}_M = 17.5$, and $\widehat{\sigma}_H = 68.5$.

TABLE 5 & FIGURE 2

4.2 Statistical and Economic Significance

To assess the relative statistical importance of probability distortions and standard risk aversion, we estimate restricted models and perform Vuong (1989) model selection tests.[26] We find that a model with probability distortions alone is "better" at the 1 percent level than a model with standard risk aversion alone. However, a likelihood ratio test rejects at the 1 percent level both (i) the null hypothesis of standard risk neutrality ($r = 0$) for Models 1a and 1b and (ii) the null hypothesis of no probability distortions ($\Omega(\mu) = \mu$) for

[24] Here we impose the restriction that $\Omega(\mu) > 0$.

[25] Figure 2 shows the estimated $\Omega(\mu)$ on the interval $(0, 0.16)$, wherein lie 98.2 percent of the predicted claim probabilities in the core sample.

[26] Vuong's (1989) test allows one to select between two nonnested models on the basis of which best fits the data. Neither model is assumed to be correctly specified. Vuong (1989) shows that testing whether one model is significantly closer to the truth (its loglikelihood value is significantly greater) than another model amounts to testing the null hypothesis that the loglikelihoods have the same expected value.

Model 1b.[27] This suggests that probability distortions and standard risk aversion both play a statistically significant role.

To give a sense of the economic significance of our estimates for r and $\Omega(\mu)$, we consider the implications for a household's maximum willingness to pay (WTP) for lower deductibles. Specifically, we consider a household's WTP to reduce its deductible from \$1000 to \$500 when the premium for coverage with a \$1000 deductible is \$200. Table 6 displays WTP for selected claim probabilities μ and several preference combinations. Column (1) displays WTP for a risk neutral household. Columns (2) through (4) display WTP for preference combinations using the estimates for r and $\Omega(\mu)$ from Model 1a. Lastly, column (5) displays WTP for a household with our estimated degree of standard risk aversion when we do not allow for probability distortions. Comparing columns (2) through (4) reveals that our estimated probability distortions have a larger economic impact than our estimated standard risk aversion, except at claim probabilities at the high end of our data, where the impacts are comparably large. Comparing columns (4) and (5) reveals that the best fit of a model which features only standard risk aversion is overly sensitive to changes in claim probability.

TABLE 6

Lastly, to give a sense of the economic significance of our estimates for the scale of choice noise (σ_L, σ_M, and σ_H), we consider the potential for such noise to affect households' deductible choices. In particular, we use Model 1a to simulate—both with and without choice noise—the deductible choices of the households in the core sample. Over 1000 iterations, we find that the households' "noisy" choices match their "noiseless" choices about half the time: 56 percent in auto collision, 43 percent in auto comprehensive, and 50 percent in home. Moreover, we find that households' "noisy" choices are within one rank (i.e., one step up or down on the menu of deductible options) of their "noiseless" choices more than four-fifths of the time: 94 percent in collision, 82 percent in auto comprehensive, and 95 percent in home. This suggests that choice noise, at the scale we estimate, is important but not dominant.[28]

4.3 Sources of Probability Distortions

There are a number of possible sources of the estimated probability distortions depicted in Figure 2. In this section, we discuss what we learn from our estimated $\Omega(\mu)$ about several potential sources.

[27] We do not perform a likelihood ratio test of the null hypothesis of no probability distortions for Model 1a because it does not nest the case $\Omega(\mu) = \mu$.

[28] After all, with extreme noise, we would find match rates that are two to three times smaller, because the probability that a household's "noisy" choice would equal any given deductible would be 20 percent in auto collision, 17 percent in auto comprehensive, and 25 percent in home.

One potential source of probability distortions is probability weighting, whereby probabilities are transformed into decision weights.[29] Under a probability weighting model, and adopting the rank-dependent approach of Quiggin (1982), the utility of deductible lottery L_d is given by

$$U(L_d) = (1 - \pi(\mu))\, u(w - p_d) + \pi(\mu)\, u(w - p_d - d)\,, \tag{4}$$

where $\pi(\mu)$ is the probability weighting function. Clearly, equation (4) is equivalent to equation (1) with $\Omega(\mu) = \pi(\mu)$. Insofar as $\Omega(\mu)$ reflects probability weighting, our estimated $\Omega(\mu)$ is striking in its resemblance to the probability weighting function originally posited by Kahneman and Tversky (1979). In particular, it is consistent with a probability weighting function that exhibits overweighting of small probabilities, exhibits mild insensitivity to changes in probabilities, and trends toward a positive intercept as μ approaches zero (though we have relatively little data for $\mu < 0.01$). By contrast, the probability weighting functions later suggested by Tversky and Kahneman (1992), Lattimore et al. (1992), and Prelec (1998)—which are commonly used in the literature (e.g., Jullien and Salanié 2000; Kliger and Levy 2009; Bruhin et al. 2010; Andrikogiannopoulou 2010)—will not fit our data well, because they trend toward a zero intercept and typically exhibit oversensitivity for probabilities less than five to ten percent.[30]

Another possible source of probability distortions is loss aversion. The original, "status quo" loss aversion proposed by Kahneman and Tversky (1979)—wherein gains and losses are defined relative to initial wealth—cannot explain aversion to risk in the context of insurance deductible choices because all outcomes are losses relative to initial wealth. More recently, however, Kőszegi and Rabin (2007) and Sydnor (2010) have suggested that a form of "rational expectations" loss aversion proposed by Kőszegi and Rabin (2006)—wherein gains and losses are defined relative to expectations about outcomes—can explain the aversion to risk manifested in insurance deductible choices. In the Appendix, we describe the Kőszegi-Rabin (KR) model of loss aversion and derive its implications for deductible lotteries.[31] Under the

[29] As we discuss in Section 7, in our data literal probability weighting is indistinguishable from systematic risk misperceptions (i.e., incorrect subjective beliefs about claim probabilities).

[30] For instance, if we impose on $\Omega(\mu)$ the one-parameter functional form proposed by Prelec (1998), we estimate Prelec's $\alpha = 0.7$. This estimate implies that $\Omega'(\mu) > 1.0$ for $\mu < 0.075$ and $\Omega'(\mu) > 1.5$ for $\mu < 0.027$.

[31] Specifically, we apply the concept of "choice-acclimating personal equilibrium" (CPE), which KR suggest is appropriate for insurance choices because the insured commits to its policy choices well in advance of the resolution of uncertainty. In addition, we explain that although the KR model contains two parameters, one (λ) that captures the degree of loss aversion and one (η) that captures the importance of gain-loss utility relative to standard utility, under CPE these parameters always appear as the product $\eta(\lambda - 1)$, which we label Λ.

KR model, the utility of deductible lottery L_d is given by

$$\begin{aligned} U(L_d) &= [(1-\mu)u(w-p_d)+\mu u(w-p_d-d)] \\ &\quad -\Lambda\,(1-\mu)\,\mu\,[u(w-p_d)-u(w-p_d-d)]\,. \end{aligned} \tag{5}$$

The first bracketed term is merely the standard expected utility of L_d. The second bracketed term reflects the expected disutility due to loss aversion, where $\Lambda \geq 0$ captures the degree of loss aversion ($\Lambda = 0$ means no loss aversion). When $\Lambda > 0$, the outcome of experiencing a claim "looms larger" than the outcome of not experiencing a claim because the former is perceived as a loss relative to the latter. Equation (5) is equivalent to equation (1) with $\Omega(\mu) = \mu + \Lambda\,(1-\mu)\,\mu$. If $\Lambda > 0$ then $\Omega(\mu) \neq \mu$. Thus, KR loss aversion can generate probability distortions. The probability distortions implied by KR loss aversion, however, are qualitatively different from the probability distortions we estimate—see panel (a) of Figure 3. In particular, the probability distortion function implied by KR loss aversion is too steep in the range of our data and trends toward a zero intercept. Hence, KR loss aversion alone cannot explain our data.

Probability distortions also can arise from disappointment aversion. In the Appendix, we describe the model of disappointment aversion proposed by Gul (1991), in which disutility arises when the outcome of a lottery is less than the certainty equivalent of the lottery.[32] Under Gul's model, the utility of deductible lottery L_d is given by

$$U(L_d) = \left(1 - \frac{\mu(1+\beta)}{1+\beta\mu}\right) u(w-p_d) + \left(\frac{\mu(1+\beta)}{1+\beta\mu}\right) u(w-p_d-d), \tag{6}$$

where $\beta \geq 0$ captures the degree of disappointment aversion ($\beta = 0$ means no disappointment aversion). When $\beta > 0$, the outcome of experiencing a claim is overweighted (relative to μ) because of the disappointment associated therewith. Equation (6) is equivalent to equation (1) with $\Omega(\mu) = \mu(1+\beta)/(1+\beta\mu)$. If $\beta > 0$ then $\Omega(\mu) \neq \mu$. Thus, Gul disappointment aversion can generate probability distortions. Again, however, the probability distortions implied by Gul disappointment aversion are qualitatively different from the probability distortions we estimate—see panel (b) of Figure 3. As before, the implied probability distortion function is too steep and trends toward a zero intercept. Hence, Gul disappointment aversion alone cannot explain our data.

Because we can "reject" the hypotheses that KR loss aversion alone or Gul disappointment aversion alone can explain our estimated probability distortions, our analysis provides

[32] In the Appendix, we explain that versions of the somewhat different approaches of Bell (1985) and Loomes and Sugden (1986) imply probability distortions identical to those implied by KR loss aversion.

evidence that probability weighting (perhaps partly reflecting risk misperceptions) is playing a key role in the households' deductible choices.

Finally, we consider a combination of sources. In the Appendix, we derive that, if households have probability weighting $\pi(\mu)$ and KR loss aversion Λ, then the utility of deductible lottery L_d is given by

$$\begin{aligned} U(L_d) &= [(1-\pi(\mu))u(w-p_d)+\pi(\mu)u(w-p_d-d)] \\ &\quad -\Lambda\,(1-\pi(\mu))\,\pi(\mu)\,[u(w-p_d)-u(w-p_d-d)]\,, \end{aligned}$$

which is equivalent to equation (1) with $\Omega(\mu)=\pi(\mu)[1+\Lambda(1-\pi(\mu))]$. From this equation, it is clear that, unless we impose strong functional form assumptions on $\pi(\mu)$, we cannot separately identify Λ and $\pi(\mu)$. Rather, the best we can do is to derive, for various values of Λ, an implied probability weighting function $\pi(\mu)$. Panel (c) of Figure 3 performs this exercise. It reinforces our conclusion that KR loss aversion alone cannot explain our estimated probability distortions, because it is clear from the figure that no value of Λ will generate an implied $\pi(\mu)$ that lies on the 45-degree line.[33]

FIGURE 3

5 Analysis with Heterogenous Preferences

In this section, we expand the model to permit heterogenous preferences. With heterogeneous preferences, it is no longer feasible to take a semi-nonparametric approach to estimating the model, due to the curse of dimensionality and to the computational burden of our estimation procedure. Hence, we now take a parametric approach to $\Omega(\mu)$. Because Models 1a, 1b, and 1c yield nearly identical results, and because it naturally constrains $\Omega(\mu)>0$, throughout this section we estimate a quadratic Chebyshev polynomial expansion of $\ln\Omega(\mu)$ (which is the best fit in Model 1a). As before, we estimate equation (3) assuming that utility is specified by equation (2) and that $\mu_{ij}=\widehat{\mu}_{ij}$, and we use combined data for all three coverages. We assume that choice noise is independent of any observed or unobserved heterogeneity in preferences, and as before we permit the scale of choice noise to vary across coverages.

We take three approaches to modeling heterogeneity in preferences. In Model 2, we allow for observed heterogeneity in r_i and $\Omega_i(\mu)$ by assuming

$$\ln r_i=\beta_r Z_i \qquad \text{and} \qquad \ln\Omega_i(\mu)=\beta_{\Omega,1}Z_i+\left(\beta_{\Omega,2}Z_i\right)\mu+\left(\beta_{\Omega,3}Z_i\right)\mu^2,$$

[33] While we focus on a combination of probability weighting and KR loss aversion, a similar analysis of a combination of probability weighting and Gul disappointment aversion yields analogous conclusiuons.

where Z_i comprises a constant plus the variables in Table 1. We estimate Model 2 via maximum likelihood, and the parameter vector to be estimated is $\theta \equiv (\beta_r, \beta_{\Omega,1}, \beta_{\Omega,2}, \beta_{\Omega,3}, \sigma_L, \sigma_M, \sigma_H)$.

In Models 3 and 4, we allow for unobserved heterogeneity in r_i and $\Omega_i(\mu)$. In particular, we assume

$$\ln r_i = \beta_r Z_i + \xi_{r,i} \qquad \text{and} \qquad \ln \Omega_i(\mu) = \beta_{\Omega,1} Z_i + \left(\beta_{\Omega,2} Z_i\right)\mu + \left(\beta_{\Omega,3} Z_i\right)\mu^2 + \xi_{\Omega,i},$$

where

$$\begin{pmatrix} \xi_{r,i} \\ \xi_{\Omega,i} \end{pmatrix} \overset{iid}{\sim} Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Phi\right), \text{ with } \Phi \equiv \begin{bmatrix} \Phi_r & \Phi_{r,\Omega} \\ \Phi_{r,\Omega} & \Phi_\Omega \end{bmatrix}.$$

In Model 3, we allow for only unobserved heterogeneity, and thus Z_i is a constant. In Model 4, we allow for both unobserved and observed heterogeneity, and Z_i comprises a constant plus the variables in Table 1. We estimate Models 3 and 4 via Markov Chain Monte Carlo (MCMC), and the parameters to be estimated are θ and Φ.[34] Details regarding the MCMC estimation procedure are set forth in the Appendix.[35] After each estimation, we use the estimates to assign fitted values of r_i and $\Omega_i(\mu)$ to each household i.[36]

Table 7 summarizes the estimates for Models 2, 3, and 4.[37] For comparison, it also restates the estimates from Model 1a. Figure 4 depicts the mean fitted value of $\Omega(\mu)$ for each model. For comparison, it also depicts the estimated $\Omega(\mu)$ in Models 1a and 1c, along with the 95 percent confidence bands for Model 1c.

TABLE 7 & FIGURE 4

The mean estimated probability distortions in Models 2, 3, and 4 are nearly identical to each other and to the estimated probability distortions in Model 1. Hence, whether we assume preferences are homogeneous or allow for observed or unobserved heterogeneity, the main message is the same: large probability distortions characterized by substantial overweighting of claim probabilities and only mild insensitivity to probability changes.

By contrast, the estimated degree of standard risk aversion is somewhat sensitive to the modeling approach. In Model 2, the mean fitted value of r is 0.00073, slightly higher than in Model 1a. In Models 3 and 4, the estimates for r are higher still—the mean fitted values are 0.00156 and 0.00147, respectively.

[34] The resulting econometric model is a mixed multinomial logit (McFadden and Train 2000). Within the literature on estimating risk preferences, a similar specification is employed by Andrikogiannopoulou (2010).

[35] The procedure closely follows Train (2009, ch. 12). The estimation was performed in MATLAB using a modified version of Train's software, "Mixed logit estimation by Bayesian methods." Convergence diagnostic tests were run using the CODA package adapted for MATLAB by James P. LeSage.

[36] In the case of Models 3 and 4, the fitted values are calculated taking into account the estimates for Φ. Specifically, $\widehat{r}_i = \exp(\widehat{\beta}_r Z_i + (\widehat{\Phi}_r/2))$ and $\widehat{\Omega}_i(\mu) = \exp(\widehat{\beta}_{\Omega,1} Z_i + (\widehat{\beta}_{\Omega,2} Z_i)\mu + (\widehat{\beta}_{\Omega,3} Z_i)\mu^2 + (\widehat{\Phi}_\Omega/2))$.

[37] The complete estimates are reported in Tables A.6, A.7, and A.8 in the Appendix.

The estimates for the scale of choice noise are similarly sensitive. Whereas the estimates for Model 2 differ only slightly from those in Model 1a, the estimates for Models 3 and 4 (though similar to each other) differ sizably from those in Model 1a—roughly speaking, $\widehat{\sigma}_L$ and $\widehat{\sigma}_M$ are forty percent lower and $\widehat{\sigma}_H$ is forty percent higher. That said, the estimates for each model display the same qualitative pattern: $\widehat{\sigma}_M < \widehat{\sigma}_L << \widehat{\sigma}_H$.

Lastly, we make two observations about the estimated variance-covariance matrix of unobserved heterogeneity. First, the variance estimates imply that there indeed is unobserved heterogeneity in preferences. Consider, for instance, Model 3 (the message is the same for Model 4, although the calculations are more involved because of the presence of observed heterogeneity). For standard risk aversion, the estimate of $\hat{\Phi}_r = 0.55$ implies that the 2.5th, 25th, 75th, and 97.5th percentiles are 0.00028, 0.00072, 0.00195, and 0.00507, respectively. For probability distortions, the estimate of $\hat{\Phi}_\Omega = 0.37$ implies that the 2.5th, 25th, 75th, and 97.5th pointwise percentiles are $0.25\overline{\Omega}(\mu)$, $0.55\overline{\Omega}(\mu)$, $1.25\overline{\Omega}(\mu)$, and $2.73\overline{\Omega}(\mu)$, where $\overline{\Omega}(\mu)$ is the mean fitted probability distortion depicted in Figure 4. This substantial unobserved heterogeneity is consistent with similar findings on standard risk aversion by Cohen and Einav (2007) and on cumulative prospect theory parameters by Andrikogiannopoulou (2010).[38] We note, however, that we find less unobserved heterogeneity in standard risk aversion than do Cohen and Einav (2007).[39] Despite this unobserved heterogeneity, our main message persists: the data is best explained by large probability distortions among the majority of households. Furthermore, our conclusions regarding the sources of probability distortions—and in particular that neither KR loss aversion alone nor Gul disappointment aversion alone can explain our estimated probability distortions—also persist.

Our second observation is that the covariance estimate implies a negative correlation between unobserved heterogeneity in r and $\Omega(\mu)$: -0.49 in both Models 3 and 4.[40] This suggests that the unexplained variation in the households' deductible choices (after controlling for premiums, claim probabilities, and observed heterogeneity) is best explained not by heterogeneity in their "overall" aversion to risk but rather by heterogeneity in their combi-

[38] It also is consistent with findings by Chiappori et al. (2012) and Gandhi and Serrano-Padial (2012), who use market-level gambling data to study, respectively, heterogeneity in preferences and heterogeneity in beliefs (which, incidentally, feature non-vanishing chances being assigned to events with vanishing probabilities).

[39] This is perhaps not surprising given that we have a second dimension of unobserved heterogeneity in preferences. Because Andrikogiannopoulou (2010) estimates a very different model—which in particular assumes no standard risk aversion and imposes a functional form for probability weighting that reflects a very different shape from ours—it is difficult to compare the magnitude of our estimates of unobserved heterogeneity with those in her paper.

[40] To be clear, Table 7 reports that the implied correlation between $\xi_{r,i}$ and $\xi_{\Omega,i}$ is -0.72. Given the log-linear specifications for r_i and $\Omega_i(\mu)$, this implies that the correlation between r_i and $\Omega_i(\mu)$ due to unobserved heterogeneity is -0.49.

nations of standard risk aversion and probability distortions. After all, if such unexplained variation in deductible choices were best explained by heterogeneity in overall aversion to risk, we would expect a positive correlation because moving r and $\Omega(\mu)$ in the same direction is the most "efficient" way of varying overall aversion to risk. However, if there were little heterogeneity in overall aversion to risk, moving r and $\Omega(\mu)$ in opposite directions would be necessary to explain such variation in deductible choices.

6 Sensitivity Analysis

Our main analysis yields a clear main message: large probability distortions characterized by substantial overweighting and mild insensitivity. Moreover, our main analysis suggests that this message is robust to different approaches to modeling heterogeneity in preferences. In this section, we further investigate the sensitivity of this message, and we find that it is robust to various other modeling assumptions. By contrast, we generally find that the estimates for standard risk aversion are more sensitive. To conserve space, we only summarize the results of the sensitivity analysis below. The complete results are available in the Appendix (Tables A.9 through A.22). In most of the sensitivity analysis, we restrict attention to Models 2 and 3. (Recall that Model 2 allows for observed heterogeneity in preferences and Model 3 allows for unobserved heterogeneity in preferences.) We do not re-estimate Model 4 because of the extreme computational burden of estimating the model with both observed and unobserved heterogeneity,[41] and also because Models 2, 3 and 4 (not to mention Model 1) all yield the same main message.

6.1 Unobserved Heterogeneity in Risk

In our main analysis, we assume that the households' subjective claim probabilities correspond to our predicted claim probabilities, which reflect only observed heterogeneity. The results of our claim rate regressions, however, imply that there is unobserved heterogeneity. In this section, we take two approaches to accounting for this unobserved heterogeneity.

In our first approach, we assume that unobserved heterogeneity in risk is not correlated with unobserved heterogeneity in preferences. With this assumption, we use the results of the claim rate regressions to assign to each household a predicted distribution of claim probabilities, $\widehat{F}(\mu)$, and then integrate over $\widehat{F}(\mu)$ to construct the likelihood function. Column (a) of Table 8 summarizes the estimates for Models 2 and 3 when we allow for unobserved

[41] Estimating Model 4 takes approximately one week (on a Dell Precision 7500 with dual XEON 5680 processors with 24GB of memory). In comparison, estimating Model 3 takes approximately one day, and estimating Model 2 takes less than an hour.

heterogeneity in risk in this way. For comparison, the table also restates the benchmark estimates. Our main message remains unchanged. The estimates for $\Omega(\mu)$ indicate similarly large probability distortions—see Figure A.1 in the Appendix. The estimates for r indicate levels of standard risk aversion that are somewhat higher than the benchmark estimates. Lastly, we note that the estimates for the scale of choice noise, as well as the estimates for Φ, are similar to the benchmark estimates.

TABLE 8

In our second approach, we allow for correlation between unobserved heterogeneity in risk and unobserved heterogeneity in preferences. In principle, we could estimate Model 3 with the addition of unobserved heterogeneity in claim probabilities and permit a flexible correlation structure among the various sources of unobserved heterogeneity. Doing so, however, would impose an undue computational burden and put a strain on identification. Instead, we estimate the following model:

$$\ln r_i = r + \xi_{r,i} \qquad \text{and} \qquad \Omega_{ij}(\mu) = a + b\widehat{\mu}_{ij}\exp(\xi_{b,ij}),$$

where $(\xi_{r,i}, \xi_{b,iL}, \xi_{b,iM}, \xi_{b,iH}) \overset{iid}{\sim} Normal(0, \Psi)$ and the parameters to be estimated are r, a, b, and Ψ. Relative to Model 3, this model imposes a more restrictive functional form on $\Omega(\mu)$ and also alters the way in which unobserved heterogeneity enters into $\Omega(\mu)$. However, it has several compensating virtues. Perhaps most important, given the way ξ_b enters the model, it can be interpreted as unobserved heterogeneity in probability distortions (as in our first approach), but it also can be interpreted as unobserved heterogeneity in subjective risk perceptions (i.e., subjective claim probabilities). Thus, the model nests—when $a = 0$ and $b = 1$—a standard expected utility model with no probability distortions but with unobserved heterogeneity in both risk and standard risk aversion (and a flexible correlation structure).[42] In addition, it allows for coverage-specific unobserved heterogeneity, and it is simple and computationally tractable.

Table A.11 in the Appendix reports our MCMC estimates for this model. For $\Omega(\mu)$, we obtain an intercept of $\widehat{a} = 0.041$ (standard error: 0.002) and a slope of $\widehat{b} = 0.864$ (standard error: 0.029). These estimates clearly reject a standard expected utility model and yield probability distortions that are quite similar to Model 3.[43] Hence, our main message remains

[42]Technically, this nesting is complete only if we further impose that $(\xi_{r,i}, \xi_{b,iL}, \xi_{b,iM}, \xi_{b,iH}) \overset{iid}{\sim} Normal(\Upsilon, \Psi)$, where $\Upsilon_j = -\Psi_{jj}/2$. If we impose this restriction, the results are very similar to those we report (which assume $\Upsilon = 0$).

[43]The somewhat lower intercept is compensated by the higher variance of the unobserved heterogeneity terms, Ψ_b, which imply a higher $E(\exp(\xi_b))$.

unchanged. The estimates for standard risk aversion and for the scale of choice noise are also similar to Model 3. The estimates for the variance-covariance matrix Ψ imply much higher variances of unobserved heterogeneity in risk than do the estimates for ϕ from the claim rate regressions, suggesting either that there is substantial unobserved heterogeneity in probability distortions or that there indeed is unobserved heterogeneity in subjective risk perceptions beyond the unobserved heterogeneity in objective risk. Interestingly, the variance estimates are quite similar for auto collision and home, but substantially higher for auto comprehensive, perhaps suggesting that households have more accurate beliefs about risk in domains in which adverse events occur more frequently. Lastly, the cross-coverage correlations of unobserved heterogeneity in risk are all high, supporting our benchmark modeling assumption that there is a household-specific component in probability distortions that is common across coverages.

Finally, we note that unobserved heterogeneity in risk creates an adverse selection problem for the insurance company. However, adverse selection from the company's perspective is irrelevant to our analysis. What matters here is to account for selection based on unobservable risk, which is precisely what we do.

6.2 Restricted Choice Noise

In the model, we use choice noise to account for observationally equivalent households choosing different deductibles, and for individual households making "inconsistent" choices across coverages. In this section, we investigate the sensitivity of our main message to this modeling choice by restricting the scale of choice noise. Column (b) of Table 8 summarizes the estimates for Models 2 and 3 when we restrict the scale of choice noise to half its estimated magnitude (i.e., we fix $\sigma_j \equiv \frac{1}{2}\widehat{\sigma}_j$). Our main message is unchanged. Indeed, the estimated probability distortions become even more pronounced—see Figure A.2 in the Appendix. At the same time, the estimate for standard risk aversion becomes extremely small.

6.3 CRRA Utility

In our analysis, we use a second-order Taylor expansion of the utility function. As a check of this approach, we consider CRRA utility, $u(w) = w^{1-\rho}/(1-\rho)$, where $\rho > 0$ is the coefficient of relative risk aversion. The CRRA family is "the most widely used parametric family for fitting utility functions to data" (Wakker 2008). With CRRA utility, equation (1) becomes

$$U(L_d) = (1 - \Omega(\mu)) \frac{(w - p_d)^{1-\rho}}{(1-\rho)} + \Omega(\mu) \frac{(w - p_d - d)^{1-\rho}}{(1-\rho)}.$$

A disadvantage of CRRA utility is that it requires wealth as an input, and moreover there surely is heterogeneity in wealth across the households in our core sample. To account for these issues, we assume that (i) wealth is proportional to home value and (ii) average wealth is $33,000 (approximately equal to 2010 U.S. per capita disposable personal income). The average home value in the core sample is approximately $191,000. Thus, we assume $w = (33/191) \times$ (home value).[44]

We re-estimate Models 2 and 3.[45] For Model 2, we assume $\ln \rho = \beta_\rho Z_i$, where Z_i comprises a constant plus the variables in Table 1. For Model 3, we assume $\ln \rho = \beta_\rho + \xi_{\rho,i}$. We otherwise proceed exactly as described in Section 5. Panel (c) of Table 8 summarizes the estimates, and Figure A.3 in the Appendix depicts the mean estimated $\Omega(\mu)$ for each model. The main message is much the same: we find large probability distortions characterized by substantial overweighting and mild insensitivity. In fact, the probability distortions become somewhat more pronounced. At the same time, the estimates for standard risk aversion become very small. The mean fitted values of ρ are 0.37 and 0.21 in Models 2 and 3, respectively. Evaluated at average wealth of $33,000, these estimates imply a coefficient of absolute risk aversion on the order of $r = 0.00001$.

6.4 Moral Hazard

Throughout our analysis, we assume that deductible choice does not influence claim risk. That is, we assume there is no deductible-related moral hazard. In this section, we assess this assumption.

There are two types of moral hazard that might operate in our setting. First, a household's deductible choice might influence its incentives to take care (ex ante moral hazard). Second, a household's deductible choice might influence its incentives to file a claim after experiencing a loss (ex post moral hazard), especially if its premium is experience rated or if the loss results in a "nil" claim (i.e., a claim that does not exceed its deductible). For either type of moral hazard, the incentive to alter behavior—i.e., take more care or file fewer claims—is stronger for households with larger deductibles. Hence, we investigate whether moral hazard is a significant issue in our data by examining whether our predicted claim probabilities change if we exclude households with high deductibles.

Specifically, we re-run our claim rate regressions using a restricted sample of the full data set in which we drop all household-coverage-year records with deductibles of $1000 or

[44] We also restrict households to have positive wealth. The results presented here restrict $w \geq \$12,000$ (i.e., if a household's implied w is less than $12,000, we assign it a wealth of $12,000). We investigated sensitivity to this cutoff, and it matters very little (because very few households are affected).

[45] To enable a global search over ρ, we scale the model by $(33,000)^\rho$. Locally, the results are indistinguishable from those obtained without re-scaling.

larger.[46] We then use the new estimates to generate revised predicted claim probabilities for all households in the core sample (including those with deductibles of $1000 or larger). Comparing the revised predicted claim rates with the benchmark predicted claim rates, we find that they are essentially indistinguishable—in each coverage, pairwise correlations exceed 0.995 and linear regressions yield intercepts less than 0.001 and coefficients of determination (R^2) greater than 0.99. Moreover, the estimates of the variance of unobserved heterogeneity in claim rates are nearly identical.[47] Not surprisingly, if we re-estimate our baseline model using the revised predicted claim probabilities, the results are virtually identical.

The foregoing analysis suggests that moral hazard is not a significant issue in our data. This is perhaps not surprising, for two reasons. First, the empirical evidence on moral hazard in auto insurance markets is mixed. (We are not aware of any empirical evidence on moral hazard in home insurance markets.). Most studies that use "positive correlation" tests of asymmetric information in auto insurance do not find evidence of a correlation between coverage and risk (e.g., Chiappori and Salanié 2000; for a recent review of the literature, see Cohen and Siegelman 2010).[48] Second, there are theoretical reasons to discount the force of moral hazard in our setting. In particular, because deductibles are small relative to the overall level of coverage, ex ante moral hazard strikes us as implausible in our setting.[49] As for ex post moral hazard, households have countervailing incentives to file claims no matter the size of the loss—under the terms of the company's policies, if a household fails to report a claimable event (especially an event that is a matter of public record—e.g., collision events typically entail police reports), it risks denial of all forms of coverage (notably liability coverage) for such event and also cancellation (or nonrenewal) of its policy.

Finally, we note that, even if our predicted claim rates are roughly correct, the possibility of nil claims could bias our results, as they violate our assumption that every claim exceeds the highest available deductible (which underlies how we define the deductible lotteries). To investigate this potential, we re-estimate Model 1a under the extreme counterfactual assump-

[46] We draw the line at the $1000 deductible and not the $500 deductible because realistically it is difficult to imagine claimable events that result in losses smaller than $500.

[47] The revised estimates are 0.22, 0.56, and 0.44 in auto collision, auto comprehensive, and home, respectively, whereas the corresponding benchmark estimates are 0.22, 0.57, and 0.45.

[48] Beginning with Abbring et al. (2003a,b), a second strand of literature tests for moral hazard in longitudinal auto insurance data using various dynamic approaches. Abbring et al. (2003b) find no evidence of moral hazard in French data. A handful of subsequent studies present some evidence of moral hazard using data from Canada and Europe. The only study of which we are aware that uses U.S. data is Israel (2004), which reports a small moral hazard effect for drivers in Illinois. Each of these studies, however, identifies a moral hazard effect with respect to either liability coverage or a composite coverage that confounds liability coverage with other coverages. None of them identifies a separate moral hazard attributable to the choice of deductible in the auto coverages we study.

[49] We note that Cohen and Einav (2007) reach the same conclusion. Furthermore, we note that the principal justification for deductibles is the insurer's administrative costs (Arrow 1963).

tion that claimable events invariably result in losses between \$500 and \$1000, specifically \$750. With this assumption, our model is unchanged except that the lottery associated with a \$1000 deductible becomes $L_{1000} \equiv (-p_{1000}, 1-\mu; -p_{1000} - 750, \mu)$. Because this change makes the \$1000 deductible more attractive, we will need more overall aversion to risk to explain households choosing smaller deductibles—i.e., r or $\Omega(\mu)$ will need to increase. It turns out that the estimates for $\Omega(\mu)$ indicate very similar probability distortions—see Figure A.4 in the Appendix. What changes is the estimate for standard risk aversion: $\widehat{r}$ increases to 0.002. This suggests that our main message is robust to the possibility of nil claims.

6.5 Additional Sensitivity Checks

In the Appendix, we report the results of several additional sensitivity checks, in which we consider: constant absolute risk aversion (CARA) utility; alternative samples of the data; restricted menus of deductible options; and alternative assumptions about the structure of choice noise. In each case, the estimates indicate probability distortions that are similar to the benchmark, further reinforcing our main message.

7 Discussion

We develop a structural model of risky choice that permits standard risk aversion and nonstandard probability distortions, and we estimate the model using data on households' deductible choices in auto and home insurance. We find that large probability distortions—characterized by substantial overweighting of small probabilities and only mild insensitivity to probability changes—play a statistically and economically significant role in explaining households' deductible choices. Our results yield important insights about the possible sources of these probability distortions. In particular, our analysis offers evidence of probability weighting, and suggests a probability weighting function that closely resembles the form originally suggested by Kahneman and Tversky (1979). In addition, we can "reject" the hypothesis that KR loss aversion alone or Gul disappointment aversion alone can explain our estimated probability distortions, though we cannot say whether they might be playing a role in conjunction with probability weighting.

Perhaps the main takeaway of the paper is that economists should pay greater attention to the question of how people evaluate risk. Prospect theory incorporates two key features: a value function that describes how people evaluate outcomes and a probability weighting function that describes how people evaluate risk. The behavioral literature, however, has focused primarily on the value function, and there has been relatively little focus on probability weighting.[50] In light of our work, as well as other recent work that reaches similar conclusions using different data and methods, it seems clear that future research on decision making under uncertainty should focus more attention on probability weighting.[51]

Another takeaway relates to Rabin's (2000) critique of expected utility theory. Rabin demonstrates that reliance on the expected utility model to explain aversion to moderate-stakes risk is problematic, because the "estimated" model would imply an implausible degree of risk aversion over large-stakes risk.[52] Indeed, when we estimate an expected utility model—which does not permit probability distortions—our estimate of standard risk aversion is "too large" in the Rabin sense. However, when we estimate our model—which permits probability distortions—there is far less standard risk aversion. This suggests that it may be possible—contrary to what some have argued—to resolve Rabin's anomaly without moving to models that impose zero standard risk aversion and use a non-standard value function to explain aversion to risk.

That said, it is worth highlighting certain limitations of our analysis. An important limitation is that, while our analysis clearly indicates that a lot of "action" lies in how people evaluate risk, it does not enable us to say whether households are engaging in probability weighting per se or whether their subjective beliefs about risk simply do not correspond to the objective probabilities.[53] In some ways, this distinction is not important, because it is irrelevant for predicting simple risky choices like those we study in this paper. But in other ways, this distinction is quite relevant. Notably, the arguments in favor of policy interventions to educate households about the risks they face have more purchase if our estimated probability distortions reflect risk misperceptions. They have less purchase, and perhaps none, if our estimated probability distortions reflect probability weighting.

Another important limitation is that our analysis relies exclusively on insurance deductible choices, and hence we urge caution when generalizing our conclusions to other domains. In particular, the vast majority of the claim probabilities we observe lie between zero and sixteen percent, and thus our analysis implies little about what probability distortions

[50] Two prominent review papers—an early paper that helped set the agenda for behavioral economics (Rabin 1998) and a recent paper that surveys the current state of empirical behavioral economics (DellaVigna 2009)—contain almost no discussion of probability weighting. The behavioral finance literature has paid more attention to probability weighting (see, e.g., Barberis and Huang 2008; Barberis 2012)

[51] Indeed, Prelec (2000) conjectured that "probability nonlinearity will eventually be recognized as a more important determinant of risk attitudes than money nonlinearity."

[52] More narrowly, Drèze (1981) and Sydnor (2010) describe how real-world insurance deductibles seem to imply "too much" risk aversion.

[53] Relatedly, although probability weighting is a natural candidate for explaining the probability distortions that we find, other mechanisms, such as ambiguity aversion, also could give rise to similar probability distortions. In fact, Fellner (1961) suggests using "distorted probabilities" to model the aversion to ambiguity in the Ellsberg (1961) paradox.

might look like outside that range. While we suspect that our main message would resonate in many domains beyond insurance deductible choices that involve similar probabilities, we hesitate to make any conjectures about settings where larger probabilities are involved.

Finally, we highlight a natural question that arises from our analysis: Are firms aware of the nature of households' risk preferences and do they react optimally to these risk preferences? Investigating this question would require significant further thought. Even if we thought that the insurance company were a risk neutral expected profit maximizer, it would be too simplistic to assume that it merely maximizes the (negative) expected value of the deductible lottery chosen by households in our setting. Optimal insurance contracts also depend on the legal restrictions imposed by regulators, the nature of competition with other insurance companies, and the nature of an insurance company's costs (for underwriting policies, servicing claims, etc.). Moreover, insurance companies care also about dynamic demand, which is something we have not considered in this paper. Hence, we view this question as beyond the scope of the present paper, but an important question for future research.

References

Abbring, J. H., P.-A. Chiappori, J. Heckman, and J. Pinquet (2003a): "Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?" *Journal of the European Economic Association*, 1, 512–521.

Abbring, J. H., P.-A. Chiappori, and J. Pinquet (2003b): "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association*, 1, 767–820.

Andrikogiannopoulou, A. (2010): "Estimating Real Preferences from a Large Panel of Real-World Betting Choices," Working Paper, Princeton University.

Arrow, K. J. (1963): "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53, 941–969.

Barberis, N. (2012): "A Model of Casino Gambling," *Management Science*, 58, 35–51.

Barberis, N. and M. Huang (2008): "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices," *American Economic Review*, 98, 2066–2100.

Barseghyan, L., J. Prince, and J. C. Teitelbaum (2011): "Are Risk Preferences Stable Across Contexts? Evidence from Insurance Data," *American Economic Review*, 101, 591–631.

Bell, D. E. (1985): "Disappointment in Decision Making under Uncertainty," *Operations Research*, 33, 1–27.

Bruhin, A., H. Fehr-Duda, and T. Epper (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78, 1375–1412.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, part B, chap. 76, 5549–5632.

Chiappori, P.-A., A. Gandhi, B. Salanié, and F. Salanié (2009): "Identifying Preferences under Risk from Discrete Choices," *American Economic Review: Papers & Proceedings*, 99, 356–362.

——— (2012): "From Aggregate Betting Data to Individual Risk Preferences," Working Paper, Columbia University.

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

Cicchetti, C. J. and J. A. Dubin (1994): "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy*, 102, 169–186.

Cohen, A. and L. Einav (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745–788.

Cohen, A. and P. Siegelman (2010): "Testing for Adverse Selection in Insurance Markets," *Journal of Risk and Insurance*, 77, 39–84.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

Drèze, J. H. (1981): "Inferring Risk Tolerance from Deductibles in Insurance Contracts," *Geneva Papers on Risk and Insurance*, 20, 48–52.

Einav, L., A. Finkelstein, I. Pascu, and M. R. Cullen (2012): "How General are Risk Preferences? Choice under Uncertainty in Different Domains," *American Economic Review*, 102, forthcoming.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.

FELLNER, W. (1961): "Distortion of Subjective Probabilities as a Reaction to Uncertainty," *Quarterly Journal of Economics*, 75, 670–689.

GANDHI, A. AND R. SERRANO-PADIAL (2012): "Does Belief Heterogeneity Explain Asset Prices: The Case of the Longshot Bias," Working Paper, University of Wisconsin-Madison.

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

ISRAEL, M. (2004): "Do We Drive More Safely When Accidents are More Expensive? Identifying Moral Hazard from Experience Rating Schemes," CSIO Working Paper 0043, Northwestern University.

JULLIEN, B. AND B. SALANIÉ (2000): "Estimating Preferences under Risk: The Case of Racetrack Bettors," *Journal of Political Economy*, 108, 503–530.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

KŐSZEGI, B. AND M. RABIN (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1166.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

KLIGER, D. AND O. LEVY (2009): "Theories of Choice under Risk: Insights from Financial Martkets," *Journal of Economic Behavior and Organization*, 71, 330–346.

LATTIMORE, P. K., J. R. BAKER, AND A. D. WITTE (1992): "The Influence of Probability on Risky Choice: A Parametric Examination," *Journal of Economic Behavior and Organization*, 17, 377–400.

LOOMES, G. AND R. SUGDEN (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies*, 53, 271–282.

MATZKIN, R. L. (1991): "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 1315–1327.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press.

——— (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski, Cambridge, MA: MIT Press.

McFadden, D. and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470.

Prelec, D. (1998): "The Probability Weighting Function," *Econometrica*, 66, 497–527.

——— (2000): "Compound Invariant Weighting Functions in Prospect Theory," in *Choices, Values, and Frames*, ed. by D. Kahneman and A. Tversky, Cambridge: Cambridge University Press.

Quiggin, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

Rabin, M. (1998): "Psychology and Economics," *Journal of Economic Literature*, 36, 11–46.

——— (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292.

Read, D., G. Loewenstein, and M. Rabin (1999): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19, 171–197.

Snowberg, E. and J. Wolfers (2010): "Explaining the Favorite-Longshot Bias: Is It Risk-Love or Misperceptions?" *Journal of Political Economy*, 118, 723–746.

Sydnor, J. (2010): "(Over)Insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177–199.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press, 2nd ed.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

Wakker, P. P. (2008): "Explaining the Characteristics of the Power (CRRA) Utility Family," *Health Economics*, 17, 1329–1344.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Table 1: Descriptive Statistics

Core Sample (4170 Households)

Variable	Mean	Std Dev	1st Pctl	99th Pctl
Driver 1 age (years)	54.5	15.4	26	84
Driver 1 female	0.37			
Driver 1 single	0.24			
Driver 1 married	0.51			
Driver 2 indicator	0.42			
Insurance score	766	113	530	987

Note: Omitted category for driver 1 marital status is divorced or separated. Insurance score is based on information contained in credit reports.

Table 2: Summary of Deductible Choices

Core Sample (4170 Households)

Deductible	Collision	Comp	Home
$50		5.2	
$100	1.0	4.1	0.9
$200	13.4	33.5	
$250	11.2	10.6	29.7
$500	67.7	43.0	51.9
$1000	6.7	3.6	15.9
$2500			1.2
$5000			0.4

Note: Values are percent of households.

Table 3: Summary of Premium Menus

Core Sample (4170 Households)

Coverage	**Mean**	**Std Dev**	**1st Pctl**	**99th Pctl**
Auto collision premium for $500 deductible	180	100	50	555
Auto comprehensive premium for $500 deductible	115	81	26	403
Home all perils premium for $500 deductible	679	519	216	2511
Cost of decreasing deductible from $500 to $250:				
Auto collision	54	31	14	169
Auto comprehensive	30	22	6	107
Home all perils	56	43	11	220
Savings from increasing deductible from $500 to $1000:				
Auto collision	41	23	11	127
Auto comprehensive	23	16	5	80
Home all perils	74	58	15	294

Note: Annual amounts in dollars.

Table 4: Predicted Claim Probabilities (Annual)

Core Sample (4170 Households)

	Collision	**Comp**	**Home**
Mean	0.069	0.021	0.084
Standard deviation	0.024	0.011	0.044
Correlations	**Collision**	**Comp**	**Home**
Auto collision	1		
Auto comprehensive	0.13	1	
Home all perils	0.28	0.21	1
Premium for coverage with $500 deductible	0.35	0.15	0.18

Note: Each of the correlations is significant at the 1 percent level.

Table 5: Model 1

Core Sample (4170 Households)

	Model 1a: Log $\Omega(\mu)$		Model 1b: $\Omega(\mu)$		Model 1c: Cubic Spline		
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**95% Bootstrap CI**	
r	0.00064 ***	0.00010	0.00063 ***	0.00004	0.00049	0.0000	0.0009
$\Omega(\mu)$: constant	-2.71 ***	0.03	0.061 ***	0.002			
$\Omega(\mu)$: linear	12.03 ***	0.30	1.186 ***	0.078			
$\Omega(\mu)$: quadratic	-35.15 ***	2.17	-2.634 ***	0.498			
σ_L	26.31 ***	1.14	26.32 ***	0.44	30.00	25.18	32.84
σ_M	17.50 ***	0.50	17.49 ***	0.69	25.20	20.88	27.80
σ_H	68.53 ***	5.76	66.89 ***	2.11	169.40	112.39	217.38

Notes: In Model 1a, we estimate a quadratic Chebyshev polynomial expansion of log $\Omega(\mu)$. In Model 1b, we estimate a quadraric Chebyshev polynomial expansion of $\Omega(\mu)$. In Model 1c, we estimate $\Omega(\mu)$ using an 11-point cubic spline on the interval (0,0.20).

*** Significant at 1 percent level.

Table 6: Economic Significance

	(1)	(2)	(3)	(4)	(5)
Standard risk aversion	**r=0**	**r=0.00064**	**r=0**	**r=0.00064**	**r=0.0129**
Probability distortions?	**No**	**No**	**Yes**	**Yes**	**No**
μ	**WTP**	**WTP**	**WTP**	**WTP**	**WTP**
0.020	10.00	14.12	41.73	57.20	33.76
0.050	25.00	34.80	55.60	75.28	75.49
0.075	37.50	51.60	67.30	90.19	104.86
0.100	50.00	68.03	77.95	103.51	130.76
0.125	62.50	84.11	86.41	113.92	154.00

Notes: WTP denotes—for a household with claim rate μ, the utility function in equation (2), and the specified utility parameters—the household's maximum willingness to pay to reduce its deductible from $1000 to $500 when the premium for coverage with a $1000 deductible is $200. Columns (3) and (4) use the probability distortion estimates from Model 1a.

Table 7: Models 2, 3, and 4

Core Sample (4170 Households)

	Model 1a	**Model 2**	**Model 3**	**Model 4**
r	0.00064	0.00073	0.00156	0.00147
Log Ω(μ): constant	-2.71	-2.73	-2.82	-2.83
Log Ω(μ): linear	12.03	12.40	10.72	11.22
Log Ω(μ): quadratic	-35.15	-35.61	-28.04	-31.67
σ_L	26.31	27.22	17.14	17.09
σ_M	17.50	17.91	10.16	10.46
σ_H	68.53	65.45	95.69	90.34
Φ_r			0.55	0.58
Φ_Ω			0.37	0.35
$\Phi_{r,\Omega}$			-0.33	-0.33
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)			-0.72	-0.72

Note: In Models 2, 3, and 4, the estimates reported for r are the mean fitted values.

Table 8: Sensitivity Analysis

Core Sample (4170 Households)

			(a)		(b)		(c)	
	Benchmark Estimates		**Unobserved Heterogeneity in Risk**		**Restricted Choice Noise**		**CRRA Utility**	
	Model 2	**Model 3**	**Model 2**	**Model 3**	**Model 2**	**Model 3**	**Model 2**	**Model 3**
Standard risk aversion	0.00073	0.00156	0.00097	0.00166	0.00005	0.00000	0.37	0.21
Log Ω(μ): constant	-2.73	-2.82	-2.73	-2.77	-2.52	-2.40	-2.51	-2.47
Log Ω(μ): linear	12.40	10.72	11.04	8.76	12.07	10.17	12.31	11.39
Log Ω(μ): quadratic	-35.61	-28.04	-21.78	-16.70	-36.61	-37.00	-39.53	-37.86
σ_L	27.22	17.14	25.93	17.00	13.61	8.57	22.79	13.21
σ_M	17.91	10.16	18.02	9.72	8.95	5.08	15.77	8.09
σ_H	65.45	95.69	66.01	105.49	32.72	47.84	40.24	42.41
Φ_r		0.55		0.42		2.77		0.27
Φ_Ω		0.37		0.37		0.25		0.18
$\Phi_{r,\Omega}$		-0.33		-0.28		0.05		-0.12
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)		-0.72		-0.72		0.07		-0.56

Notes: The estimates reported for standard risk aversion are the mean fitted coefficients of absolute risk aversion (r), except in the case of CRRA Utility, in which case they are the mean fitted coefficients of relative risk aversion (ρ). In the case of Restricted Choise Noise, σ_L, σ_M, and σ_H are by construction.

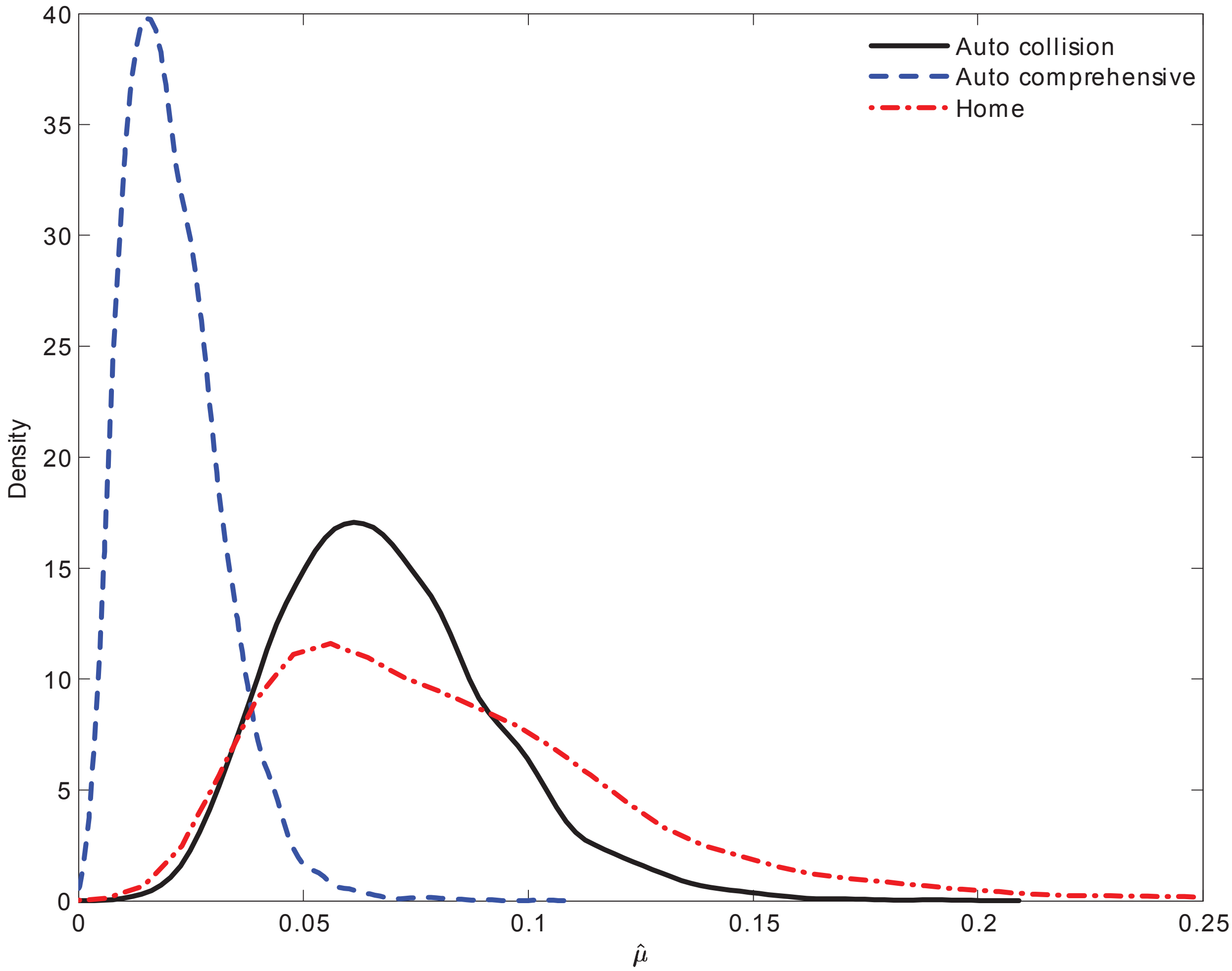


Figure 1: Empirical Density Functions for Predicted Claim Probabilities

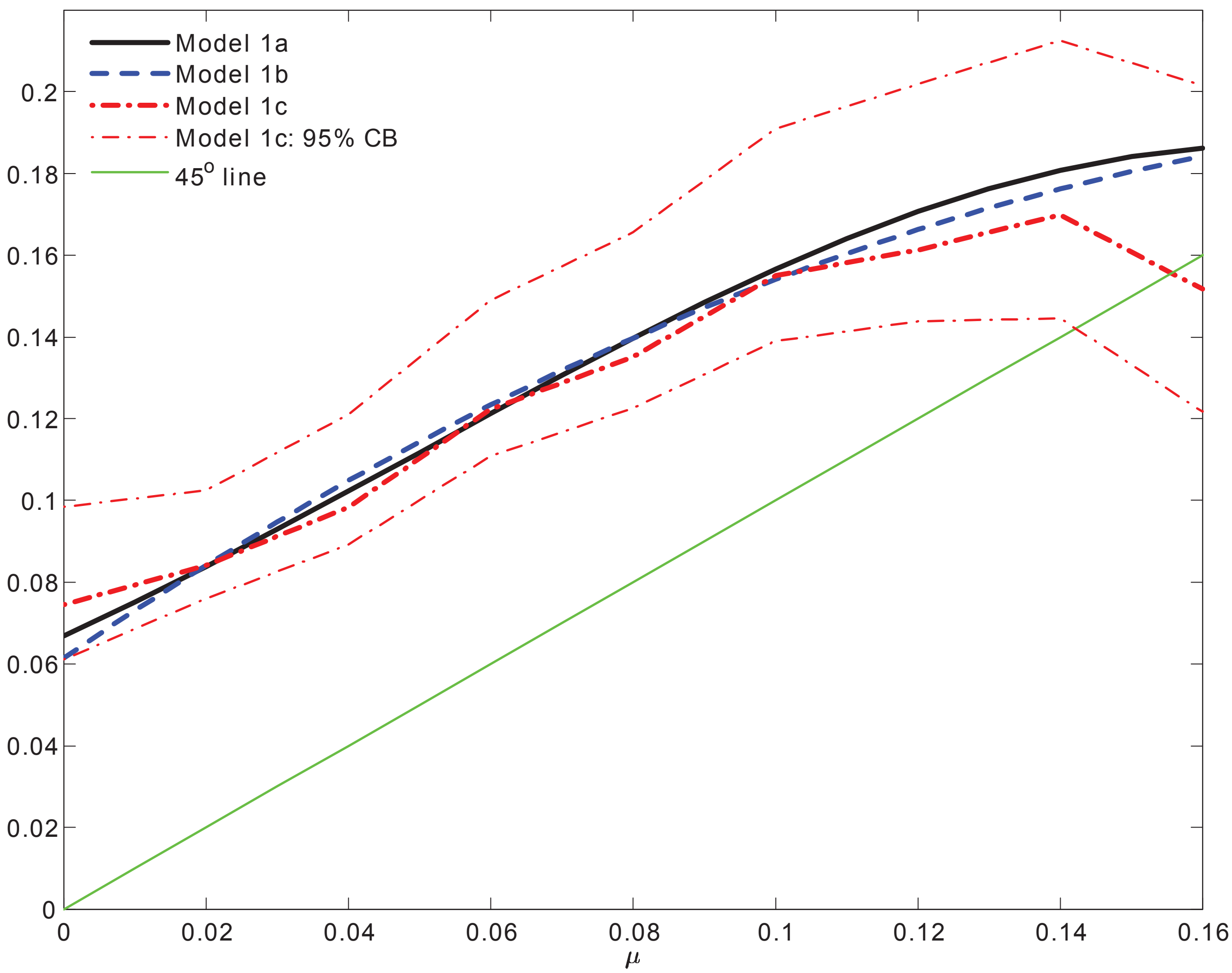


Figure 2: Estimated $\Omega(\mu)$ – Model 1

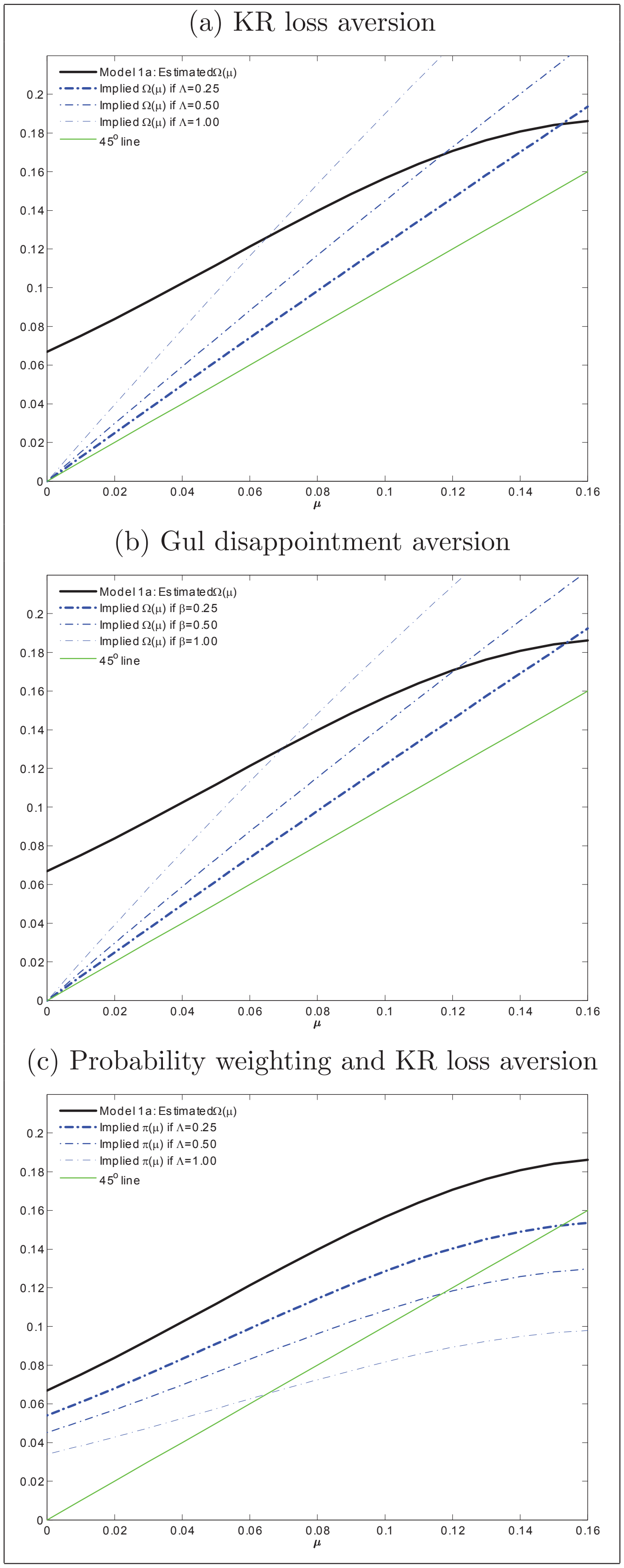


Figure 3: Sources of Probability Distortions

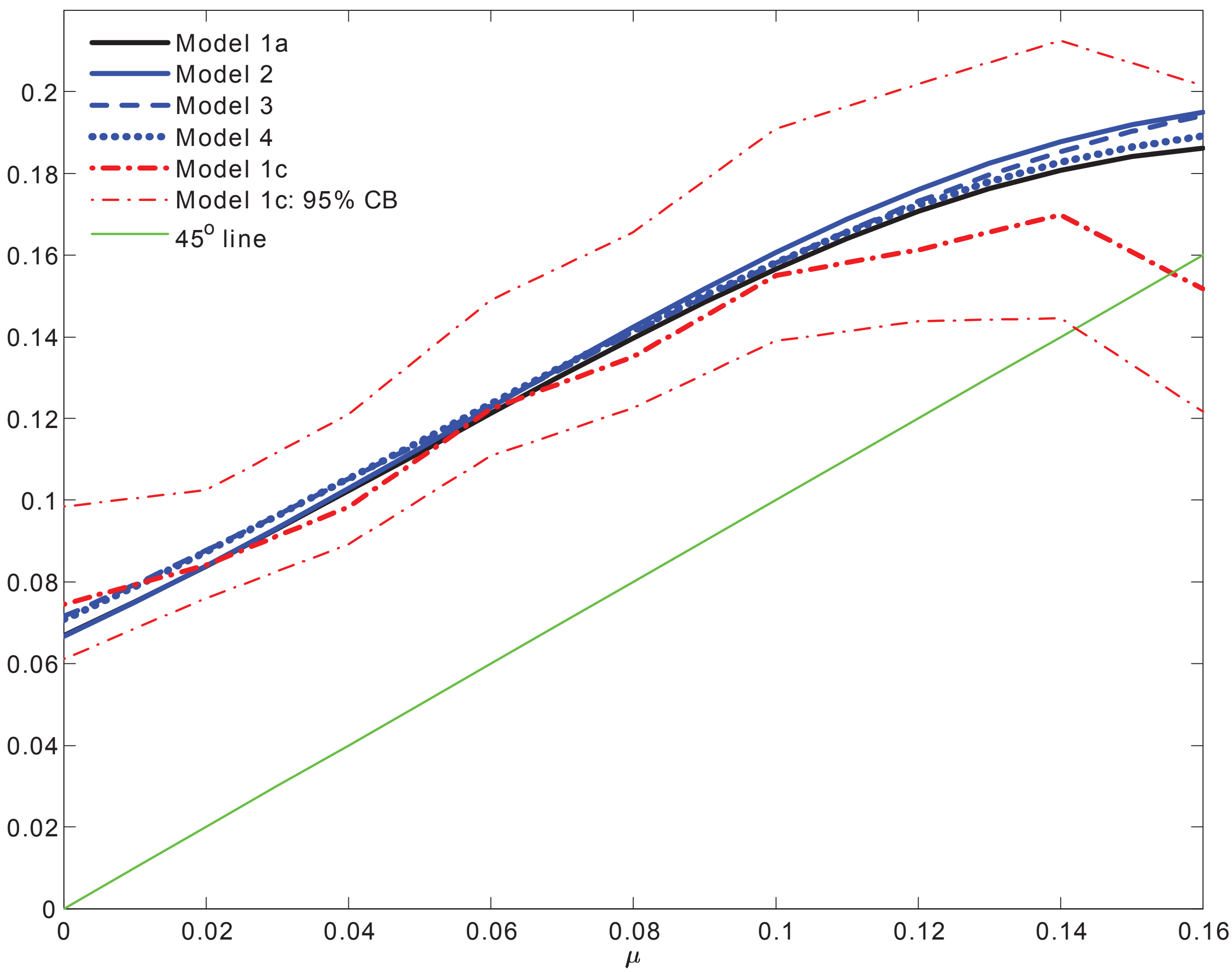


Figure 4: Mean Estimated $\Omega(\mu)$ – Models 2, 3, and 4

Appendix (Not for Publication)
to
The Nature of Risk Preferences: Evidence from Insurance Choices

Levon Barseghyan
Cornell University

Francesca Molinari
Cornell University

Ted O'Donoghue
Cornell University

Joshua C. Teitelbaum
Georgetown University

November 7, 2012

Contents

A Coverage Descriptions

Auto collision coverage pays for damage to the insured vehicle caused by a collision with another vehicle or object, without regard to fault. *Auto comprehensive* coverage pays for damage to the insured vehicle from all other causes (e.g., theft, fire, flood, windstorm, glass breakage, vandalism, hitting or being hit by an animal, or by falling or flying objects), without regard to fault. If the insured vehicle is stolen, auto comprehensive coverage also provides a certain amount per day for transportation expenses (e.g., rental car or public transportation). *Home all perils* coverage pays for damage to the insured home from all causes (e.g., fire, windstorm, hail, tornadoes, vandalism, or smoke damage), except those that are specifically excluded (e.g., flood, earthquake, or war). For simplicity, we often refer to home all perils merely as *home*.

B Identification

In this section, we prove Properties 1, 2, and 3 from Section 3.3.1. Take any three deductible options $a, b, c \in \mathcal{D}$, with $a > b > c$, and consider a household with premium p_a for deductible a and claim probability μ. The household's r and $\Omega(\mu)$ determine the premium $\widetilde{p}_b$ that makes the household indifferent between deductibles a and b, as well as the premium $\widetilde{p}_c$ that makes the household indifferent between deductibles a and c. Notice that $\widetilde{p}_b - p_a$ reflects the household's maximum willingness to pay (WTP) to reduce its deductible from a to b, and $\widetilde{p}_c - \widetilde{p}_b$ reflects the household's additional WTP to reduce its deductible from b to c. In what follows, we simplify notation by suppressing the explicit dependence of $\widetilde{p}_b$ and $\widetilde{p}_c$ on p_a, μ, r, and $\Omega(\mu)$. We also suppress the argument of Ω. In addition, let L_a denote the deductible lottery associated with deductible a at premium p_a.

Recall equation (2) from Section 3.1:

$$U(L_d) = -\left[p_d + \Omega d\right] - \frac{r}{2}\left[(1-\Omega)\left(p_d\right)^2 + \Omega\left(p_d + d\right)^2\right].$$

Define $p(x)$ as the premium for deductible x such that the household is indifferent between the resulting lottery and lottery L_a. Hence, $p(b) = \widetilde{p}_b$ and $p(c) = \widetilde{p}_c$. Applying equation (2), $p(x)$ is defined by each of the following equations (both of which we use below):[1]

$$-p(x) - \Omega x - \frac{r}{2}\left[(1-\Omega)(p(x))^2 + \Omega(p(x)+x)^2\right] = U(L_a) \qquad \text{(A.1)}$$

[1] These equations are equivalent, where the latter merely expands $U(L_a)$ and rearranges terms.

$$(p(x) - p_a) - \Omega\,(a - x) + \frac{r}{2}(p(x)^2 - p_a^2) + \Omega\frac{r}{2}\left\{(x^2 - a^2) + 2(p(x)x - p_a a)\right\} = 0. \quad \text{(A.2)}$$

For the proofs below, it is useful to define W and V as

$$\begin{aligned} W(p, x, r, \Omega) &\equiv -p - \Omega x - \frac{r}{2}\left[(1-\Omega)(p)^2 + \Omega(p+x)^2\right] \\ V(p, x, r, \Omega) &\equiv (p - p_a) - \Omega\,(a - x) + \frac{r}{2}(p^2 - p_a^2) + \Omega\frac{r}{2}\left\{(x^2 - a^2) + 2(px - p_a a)\right\}. \end{aligned}$$

Our first lemma establishes that $p(x)$ is well behaved.

Lemma 1. *For any $r \geq 0$, $\Omega \in (0, 1)$, $p_a > 0$, and $x \leq a$, $p(x)$ is a continuous and differentiable function with $dp/dx < 0$ (and thus $p(c) > p(b) > p_a$).*

It is straightforward to derive that W is twice differentiable and satisfies the conditions of the implicit function theorem. Thus $p(x)$ is a continuous and differentiable function, and

$$\frac{dp}{dx} = \frac{-\frac{\partial W}{\partial x}}{\frac{\partial W}{\partial p}} = \frac{-\Omega\,[1 + rp(x) + rx)]}{1 + rp(x) + r\Omega x} < 0.$$

Our second lemma states that standard risk aversion implies that a household's WTP to reduce its deductible is strictly greater than the expected reduction in the deductible paid (evaluated at a claim probability of Ω).

Lemma 2. *For any $x' < x \leq a$, if $r = 0$ then $p(x') - p(x) = \Omega(x - x')$, and if $r > 0$ then $p(x') - p(x) > \Omega(x - x')$.*

Proof. The result for $r = 0$ is straightforward. For $r > 0$, define $\tilde{V}$ as

$$\tilde{V}(p, x', r, \Omega) \equiv [p - p(x)] - \Omega\,(x - x') + \frac{r}{2}(p^2 - p(x)^2) + \Omega\frac{r}{2}\left\{(x'^2 - x^2) + 2(px' - p(x)x)\right\},$$

in which case $p(x')$ is defined by $\tilde{V}(p(x'), x', r, \Omega) = 0$. Note that $p = \Omega(x - x') + p(x)$ implies

$$\begin{aligned} \tilde{V}(p, x', r, \Omega) &\equiv [(\Omega(x - x') + p(x)) - p(x)] - \Omega\,(x - x') + \frac{r}{2}((\Omega(x - x') + p(x))^2 - (p(x))^2) \\ &\quad + \Omega\frac{r}{2}\left\{(x'^2 - x^2) + 2((\Omega(x - x') + p(x))x' - p(x)x)\right\} \\ &= \frac{r}{2}(\Omega^2(x - x')^2 + 2p(x)\Omega(x - x')) \\ &\quad + \Omega\frac{r}{2}\left\{(x'^2 - x^2) + 2(\Omega(x - x')x' + p(x)(x' - x))\right\} \\ &= \frac{r}{2}\Omega(x - x')[\Omega(x + x') - (x + x')] < 0. \end{aligned}$$

Since $\partial\tilde{V}/\partial p = 1 + rp + \Omega r x' > 0$, it follows that $p(x') > \Omega(x - x') + p(x)$. □

Property 1 establishes the relationship between the magnitude of willingness to pay and risk preferences.

Property 1. *For any $x < a$, $p(x)$ is strictly increasing in r and Ω.*

Proof. By implicit function theorem:

$$\frac{\partial p(x)}{\partial r} = \frac{-\frac{\partial V}{\partial r}}{\frac{\partial V}{\partial p}} \qquad \text{and} \qquad \frac{\partial p(x)}{\partial \Omega} = \frac{-\frac{\partial V}{\partial \Omega}}{\frac{\partial V}{\partial p}}.$$

Note that

$$\frac{\partial V}{\partial r} = -\frac{1}{r}[(p(x) - p_a) - \Omega\,(a - x)] < 0,$$

where the equality uses equation (A.2) and the inequality follows from Lemma 2. Note further that

$$\frac{\partial V}{\partial \Omega} = -\frac{1}{\Omega}[(p(x) - p_a) + \frac{r}{2}(p(x)^2 - p_a^2)] < 0,$$

where the equality uses equation (A.2) and the inequality follows from Lemma 1. Finally, given that $\partial V/\partial p = 1 + rp + \Omega rx > 0$, it follows that $\frac{\partial p(x)}{\partial r} > 0$ and $\frac{\partial p(x)}{\partial \Omega} > 0$. □

We next establish Property 2, which shows that a risk averse household's WTP to avoid an incremental loss depends positively on the magnitude of the absolute loss.

Property 2. *For any fixed $\Omega(\mu)$, $r = 0$ implies $\frac{p(b)-p_a}{p(c)-p(b)} = \frac{a-b}{b-c}$, and the ratio $\frac{p(b)-p_a}{p(c)-p(b)}$ is strictly increasing in r.*

Proof. The result for $r = 0$ is straightforward. As a preliminary step in proving the second part, we first prove that, for any $r > 0$, $\frac{p(b)-p_a}{p(c)-p(b)} > \frac{a-b}{b-c}$. From Lemma 1, $p(x)$ is continuous and differentiable, and thus

$$p(b) - p_a = \int_b^a \left(-\frac{dp}{dx}\right) dx \qquad \text{and} \qquad p(c) - p(b) = \int_c^b \left(-\frac{dp}{dx}\right) dx.$$

From the proof of Lemma 1, $-\frac{dp}{dx} = \frac{\Omega[1+rp(x)+rx]}{1+rp(x)+r\Omega x} > 0$, and thus

$$\frac{d\left[-\frac{dp}{dx}\right]}{dx} = \Omega\frac{r(1-\Omega)(1+rp-rx\frac{dp}{dx})}{[1+rp(x)+r\Omega x]^2} > 0.$$

In words, $-\frac{dp}{dx}$ reflects the household's marginal willingness to pay to reduce its deductible, and $-\frac{dp}{dx} > 0$ reflects that a household is indeed willing to pay a higher premium to reduce its deductible. More importantly, $d\left[-\frac{dp}{dx}\right]/dx > 0$ reflects that the larger is its deductible, the larger is the household's marginal willingness to pay to reduce that deductible (or equivalently the smaller is its deductible, the smaller is the household's marginal willingness to pay to reduce that deductible). Finally, $d\left[-\frac{dp}{dx}\right]/dx > 0$ implies

$$\begin{aligned}
p(b) - p_a &= \int_b^a \left(-\frac{dp}{dx}\right) dx > (a-b)\left(-\left.\frac{dp}{dx}\right|_{x=b}\right) \\
p(c) - p(b) &= \int_c^b \left(-\frac{dp}{dx}\right) dx < (b-c)\left(-\left.\frac{dp}{dx}\right|_{x=b}\right)
\end{aligned}$$

which together imply $\frac{p(b)-p_a}{p(c)-p(b)} > \frac{a-b}{b-c}$. With this result in hand, we now prove that $\frac{p(b)-p_a}{p(c)-p(b)}$ is strictly increasing in r. Note that $d\left(\frac{p(b)-p_a}{p(c)-p(b)}\right)/dr > 0$ if and only if $\frac{1}{p(b)-p_a}\frac{\partial p(b)}{\partial r} > \frac{1}{p(c)-p_a}\frac{\partial p(c)}{\partial r}$. Applying $\frac{\partial p(x)}{\partial r}$ from the proof of Property 1,

$$\begin{aligned}
\frac{1}{p(b)-p_a}\frac{\partial p(b)}{\partial r} &= \frac{1}{p(b)-p_a}\frac{\frac{1}{r}\left[(p(b)-p_a) - \Omega(a-b)\right]}{1+rp(b)+\Omega rb}, \\
\frac{1}{p(c)-p_a}\frac{\partial p(c)}{\partial r} &= \frac{1}{p(c)-p_a}\frac{\frac{1}{r}\left[(p(c)-p_a) - \Omega(a-c)\right]}{1+rp(c)+\Omega rc}.
\end{aligned}$$

We have

$$\frac{1}{p(b)-p_a}\left[(p(b)-p_a) - \Omega(a-b)\right] > \frac{1}{p(c)-p_a}\left[(p(c)-p_a) - \Omega(a-c)\right],$$

because

$$(p(c)-p_a)\left[(p(b)-p_a) - \Omega(a-b)\right] > (p(b)-p_a)\left[(p(c)-p_a) - \Omega(a-c)\right]$$

$$\iff \frac{(p(b)-p_a)}{a-b} > \frac{(p(c)-p_a)}{a-c}$$

where the last inequality follows from the result above—because $\frac{(p(c)-p_a)}{a-c}$ is a convex combination of $\frac{(p(b)-p_a)}{a-b}$ and $\frac{(p(c)-p(b))}{b-c}$. Finally, we have $1 + rp(b) + \Omega rb < 1 + rp(c) + \Omega rc$, because

$$rp(b) + \Omega rb < rp(c) + \Omega rc$$

$$\iff \Omega(b-c) < p(c) - p(b),$$

where the last inequality follows from Lemma 2. The result follows. □

We conclude with the key property for identification, which establishes that different pairs of r and Ω have different implications for willingness to pay (provided there are at least three deductible options on the menu).

Property 3. *If $p(b) - p_a$ is the same for $(r, \Omega(\mu))$ and $(r', \Omega(\mu)')$ with $r < r'$ (and thus $\Omega(\mu) > \Omega(\mu)'$), then $p(c) - p(b)$ is larger for $(r, \Omega(\mu))$ than for $(r', \Omega(\mu)')$.*

Proof. For a fixed $p(b)$, define $\Omega^b(r)$ by $V(p(b), b, r, \Omega^b(r)) = 0$, so that any pair $(r, \Omega^b(r))$ yields the same $p(b)$ and thus the same $p(b) - p_a$. Then

$$\frac{d\Omega^b}{dr} = \frac{-\frac{\partial V}{\partial r}}{\frac{\partial V}{\partial \Omega}} = -\frac{\frac{1}{r}\left[(p(b) - p_a) - \Omega^b(r)(a-b)\right]}{\frac{1}{\Omega^b(r)}\left[(p(b) - p_a) + \frac{r}{2}\left(p(b)^2 - p_a^2\right)\right]}.$$

Next, define $\check{p}_c(r)$ by $V(\check{p}_c(r), c, r, \Omega^b(r)) = 0$, so that $\check{p}_c(r)$ is the $p(c)$ associated with pair $(r, \Omega^b(r))$. The goal is to show that $d\check{p}_c(r)/dr < 0$, from which the result follows. Differentiating $V(\check{p}_c(r), c, r, \Omega^b(r)) = 0$ yields

$$\frac{d\left[V(\check{p}_c(r), c, r, \Omega^b(r))\right]}{dr} = \frac{\partial V}{\partial p}\frac{d\check{p}_c(r)}{dr} + \frac{\partial V}{\partial r} + \frac{\partial V}{\partial \Omega}\frac{d\Omega^b}{dr} = 0.$$

Note that

$$\begin{aligned}\frac{\partial V}{\partial r} + \frac{\partial V}{\partial \Omega}\frac{d\Omega^b}{dr} &= -\frac{1}{r}\left[(\check{p}_c(r) - p_a) - \Omega^b(r)(a-c)\right] \\ &\quad + \frac{1}{\Omega^b(r)}\left[(\check{p}_c(r) - p_a) + \frac{r}{2}(\check{p}_c(r)^2 - p_a^2)\right]\frac{\frac{1}{r}\left[(p(b) - p_a) - \Omega^b(r)(a-b)\right]}{\frac{1}{\Omega^b(r)}\left[(p(b) - p_a) + \frac{r}{2}\left(p(b)^2 - p_a^2\right)\right]}.\end{aligned}$$

We have

$$\left[(\check{p}_c(r) - p_a) - \Omega^b(r)(a-c)\right] < \left(\frac{\check{p}_c(r) - p_a}{p(b) - p_a}\right)\left[(p(b) - p_a) - \Omega^b(r)(a-b)\right]$$

as in the proof of Property 2. In addition,

$$\frac{\left[(\check{p}_c(r) - p_a) + \frac{r}{2}(\check{p}_c(r)^2 - p_a^2)\right]}{\left[(p(b) - p_a) + \frac{r}{2}\left(p(b)^2 - p_a^2\right)\right]} > \left(\frac{\check{p}_c(r) - p_a}{p(b) - p_a}\right)$$

because

$$\begin{aligned}&(p(b) - p_a)(\check{p}_c(r)^2 - p_a^2) > (\check{p}_c(r) - p_a)\left(p(b)^2 - p_a^2\right) \\ \iff &(p(b) - p_a)(\check{p}_c(r) - p_a)(\check{p}_c(r) + p_a) > (\check{p}_c(r) - p_a)(p(b) - p_a)\left(p(b) + p_a\right) \\ \iff &\check{p}_c(r) > p(b),\end{aligned}$$

where the last inequality follows from Lemma 1. Together, these imply $\frac{\partial V}{\partial r} + \frac{\partial V}{\partial \Omega}\frac{d\Omega^b}{dr} > 0$, and therefore $\frac{\partial V}{\partial p}\frac{d\check{p}_c(r)}{dr} < 0$. Hence, $\frac{\partial V}{\partial p} > 0$ implies $\frac{d\check{p}_c(r)}{dr} < 0$, and the result follows. □

C Sources of Probability Distortions

Throughout our main analysis, we assume that utility of deductible lottery L_d is given by

$$U(L_d) = (1 - \Omega(\mu))u(w - p_d) + \Omega(\mu)u(w - p_d - d), \tag{A.3}$$

where $\Omega(\mu)$ reflects probability distortions. As we discuss in Section 4.3, there are a number of possible sources of probability distortions. In this section, we describe the details of models that can generate the probability distortions in equation (A.3).

C.1 Probability Weighting

One potential source of probability distortions is probability weighting, whereby probabilities are transformed into decision weights. Under a probability weighting model, and adopting the rank-dependent approach of Quiggin (1982), the utility of deductible lottery L_d is

$$U(L_d) = (1 - \pi(\mu))u(w - p_d) + \pi(\mu)u(w - p_d - d), \tag{A.4}$$

where $\pi(\mu)$ is the probability weighting function. Clearly, equation (A.4) is equivalent to equation (A.3) with $\Omega(\mu) = \pi(\mu)$.

Over the years, several functional forms for $\pi(\mu)$ have been proposed. A seminal paper in the literature is Kahneman and Tversky (1979), which proposes that the probability weighting function should exhibit (i) overweighting of small probabilities, (ii) underweighting of large probabilities, (iii) some insensitivity to probability changes (slope less than one), and (iv) discontinuities at $\mu = 0$ and $\mu = 1$. Later papers in the literature suggest functional forms that eliminate feature (iv) (e.g., Tversky and Kahneman 1992; Lattimore et al. 1992; Prelec 1998). These all entail an oversensitivity to probability changes (slope greater than one) for μ close to zero or one, where "close" typically (for the authors' suggested parameter values) includes the bulk of the claim probabilities in our data—e.g., using the one-parameter functional form proposed by Prelec (1998) and his suggested parameter value of $\alpha = 0.65$, $\pi'(\mu) > 1.0$ for $\mu < 0.069$, $\pi'(\mu) > 1.5$ for $\mu < 0.028$, and $\pi'(\mu) > 2.0$ for $\mu < 0.015$. As we discuss in Section 4.3, insofar as our estimated $\Omega(\mu)$ reflects probability weighting, it is more in line with the function originally posited by Kahneman and Tversky (1979).

C.2 Kőszegi-Rabin Loss Aversion

Another possible source of probability distortions is loss aversion. The original, "status quo" loss aversion proposed by Kahneman and Tversky (1979)—wherein gains and losses are defined relative to initial wealth—cannot explain aversion to risk in the context of insurance deductible choices because all outcomes are losses relative to initial wealth. More recently, however, Kőszegi and Rabin (2007) and Sydnor (2010) have suggested that a form of "rational expectations" loss aversion proposed by Kőszegi and Rabin (2006)—wherein gains and losses are defined relative to expectations about outcomes—can explain the aversion to risk manifested in insurance deductible choices.

In the Kőszegi-Rabin (KR) model, the utility from choosing lottery $Y \equiv (y_n, q_n)_{n=1}^{N}$ given a reference lottery $\tilde{Y} \equiv (\tilde{y}_m, \tilde{q}_m)_{m=1}^{M}$ is

$$V(Y|\tilde{Y}) \equiv \sum_{n=1}^{N}\sum_{m=1}^{M} q_n \tilde{q}_m \left[u(y_n) + v(y_n|\tilde{y}_m)\right].$$

The function u represents standard "intrinsic" utility defined over final wealth states, just as in the expected utility model. The function v represents "gain-loss" utility that results from experiencing gains or losses relative to the reference point. For v, KR use

$$v(y|\tilde{y}) = \begin{cases} \eta\left[u(y) - u(\tilde{y})\right] & \text{if } u(y) > u(\tilde{y}) \\ \eta\lambda\left[u(y) - u(\tilde{y})\right] & \text{if } u(y) \leq u(\tilde{y}) \end{cases}.$$

In this formulation, the magnitude of gain-loss utility is determined by the intrinsic utility gain or loss relative to consuming the reference point. Moreover, gain-loss utility takes a two-part linear form, where $\eta \geq 0$ captures the importance of gain-loss utility relative to intrinsic utility and $\lambda \geq 1$ captures loss aversion. The model reduces to expected utility when $\eta = 0$ or $\lambda = 1$.

KR propose that the reference lottery equals recent expectations about outcomes—i.e., if a household expects to face lottery $\tilde{Y}$, then its reference lottery becomes $\tilde{Y}$. However, because situations vary in terms of when a household deliberates about its choices and when it commits to its choices, KR offer a number of solution concepts for the determination of the reference lottery. For insurance applications, KR suggest a "choice-acclimating personal equilibrium" (CPE). Formally:

Definition (CPE)**.** *Given a choice set $\mathcal{Y}$, a lottery $Y \in \mathcal{Y}$ is a* choice-acclimating personal equilibrium *if for all $Y' \in \mathcal{Y}$, $V(Y|Y) \geq V(Y'|Y')$.*

In a CPE, a household's reference lottery corresponds to its choice. KR argue that CPE is appropriate in situations where the household commits to a choice well in advance of the resolution of uncertainty, and thus it knows that by the time the uncertainty is resolved and it experiences utility, it will have become accustomed to its choice and hence expect the lottery induced by its choice. In particular, KR suggest that CPE is the appropriate solution concept for insurance applications.

Under the KR model with CPE, the utility to the household from choosing deductible lottery $L_d = (-p_d, 1-\mu; -p_d - d, \mu)$ is

$$\begin{aligned} U(L_d) = V(L_d|L_d) &= (1-\mu)u(w-p_d) + \mu u(w-p_d-d) \\ &\quad -\eta(\lambda-1)(1-\mu)\mu[u(w-p_d) - u(w-p_d-d)]. \end{aligned} \tag{A.5}$$

From equation (A.5), it is clear that one can not separately identify the parameters η and λ, and thus we focus on the product $\eta(\lambda-1) \equiv \Lambda$. Substituting Λ into equation (A.5) yields equation (5) in Section 4.3.

C.3 Gul Disappointment Aversion

Probability distortions also can arise from disappointment aversion. The concept of disappointment aversion was proposed by Bell (1985) and further developed by Loomes and Sugden (1986) and Gul (1991). The basic idea is that a person is disappointed (or elated) if the outcome of a lottery is worse (or better) than "expected." The approaches differ in terms of the nature of the disutility from disappointment, and in terms of the definition of what is "expected."

Here, we follow the approach of Gul (1991), in which disutility arises when the outcome of a lottery is less than the certainty equivalent for the lottery. For deductible lotteries, an intuitive way to express this model is that $U(L_d)$ is the $u(z)$ such that

$$u(z) = [(1-\mu)u(w-p_d) + \mu u(w-p_d-d)] - \beta\mu\left[u(z) - u(w-p_d-d)\right].$$

In this formulation, z represents the certainty equivalent. The first bracketed term is the standard expected utility of L_d. The second term reflects the expected disutility from disappointment that arises when the outcome is less than the certainty equivalent (which occurs in the event of a claim). The parameter β captures the magnitude of disappointment aversion, where the model reduces to expected utility for $\beta = 0$. We can rearrange this equation to yield

$$U(L_d) = \left(1 - \frac{\mu(1+\beta)}{1+\beta\mu}\right) u(w-p_d) + \left(\frac{\mu(1+\beta)}{1+\beta\mu}\right) u(w-p_d-d), \tag{A.6}$$

which is equivalent to equation (6) in Section 4.3.[2]

As illustrated in Figure 3 in Section 4.3, Gul disappointment aversion and KR loss aversion with CPE generate qualitatively similar probability distortions for claim probabilities in the range of our data. This follows from the fact that these models are quite similar to each other. Both models assume that a household experiences a form of "gain-loss utility" that depends on how its realized outcome compares to a reference point which is determined by the household's choice. The models differ only in the nature of the reference point associated with each choice—in Gul disappointment aversion the reference point is the certainty equivalent of the chosen lottery, whereas in KR loss aversion it is the chosen lottery itself. Bell (1985) and Loomes and Sugden (1986) differ from Gul (1991) in (effectively) assuming that the reference point is the standard expected utility of the chosen lottery. Combining that formulation with a two-part linear disappointment/elation function would yield a model that is equivalent to KR loss aversion with CPE.

C.4 Combinations of Sources

Finally, we consider combinations of sources. In particular, we focus on combining a probability weighting function $\pi(\mu)$ with either KR loss aversion or Gul disappointment aversion.

When we combine probability weighting and KR loss aversion, an issue arises: Given a reference lottery $\tilde{Y} \equiv (\tilde{y}_m, \tilde{q}_m)_{m=1}^M$, should the utility comparisons in $V(Y|\tilde{Y})$ use the probabilities $\tilde{q}_m$ for the comparison weights, or should they use the decision weights $\pi(\tilde{q}_m)$? KR offer no guidance on this modeling choice, as they abstract from nonlinear decision weights. To our minds, it seems natural to assume that households treat the chosen lottery and the reference lottery symmetrically. Accordingly, we assume that the decision weights are the same for the chosen lottery and the reference lottery. Under this assumption, we can rewrite equation (A.5) as

$$\begin{aligned} U(L_d) = V(L_d|L_d) \;\; &= \;\; (1-\pi(\mu))u(w-p_d) + \pi(\mu)u(w-p_d-d) \\ &\quad -\Lambda(1-\pi(\mu))\pi(\mu)[u(w-p_d) - u(w-p_d-d)], \end{aligned}$$

which is equivalent to equation (A.3) with $\Omega(\mu) = \pi(\mu)[1+\Lambda(1-\pi(\mu))]$. From this equation, it is clear that, unless we impose strong functional form assumptions on $\pi(\mu)$, we cannot separately identify Λ and $\pi(\mu)$. Rather, the best we can do is to derive, for various values of Λ, an implied probability weighting function $\pi(\mu)$. Panel (c) of Figure 3 in Section 4.3 performs this exercise.

[2] Note that equation (A.6) is equivalent to the equation at the top of page 677 in Gul (1991).

Combining probability weighting and Gul disappointment aversion is more straightforward, as we merely replace μ with $\pi(\mu)$ in equation (A.6). With this substitution, equation (A.6) is equivalent to equation (A.3) with $\Omega(\mu) = \pi(\mu)(1+\beta)/(1+\beta\pi(\mu))$. From this it is clear that, unless we impose strong functional form assumptions on $\pi(\mu)$, we cannot separately identify β and $\pi(\mu)$. Rather, the best we can do is to derive, for various values of β, an implied probability weighting function $\pi(\mu)$. The resulting figure would be very similar to panel (c) of Figure 3 in Section 4.3.

D MCMC Procedure: Hierarchical Bayes for Mixed Logit

In Models 3 and 4, we allow for unobserved heterogeneity in r_i and $\Omega_i(\mu)$. In particular, we assume

$$\ln r_i = \beta_r Z_i + \xi_{r,i} \qquad \text{and} \qquad \ln \Omega_i(\mu) = \beta_{\Omega,1} Z_i + \left(\beta_{\Omega,2} Z_i\right)\mu + \left(\beta_{\Omega,3} Z_i\right)\mu^2 + \xi_{\Omega,i},$$

where $\left(\xi_{r,i}, \xi_{\Omega,i}\right) \overset{iid}{\sim} N(0, \Phi)$. The utility from deductible $d \in \mathcal{D}$ is given by

$$\mathcal{U}(d) \equiv U(L_d) + \varepsilon_d,$$

where ε_d follows a type 1 extreme value distribution with scale parameter σ. Assuming $U(L_d)$ is specified by equation (2) in Section 3.1, we have

$$\mathcal{U}(d) = -\left[p_d + \Omega(\mu) d\right] - \frac{r}{2}\left[\left(1-\Omega(\mu)\right)(p_d)^2 + \Omega(\mu)\left(p_d + d\right)^2\right] + \varepsilon_d,$$

which can be re-written as

$$\mathcal{U}(d) = -p_d - \Omega(\mu) d - \frac{r}{2}(p_d)^2 - \frac{r}{2}\Omega(\mu)\left[(p_d+d)^2 - (p_d)^2\right] + \varepsilon_d. \tag{A.7}$$

Hence, a household i chooses deductible d in coverage j when $\mathcal{U}_{ij}(d) > \mathcal{U}_{ij}(d')$ for all $d' \neq d$, and thus the probability that household i chooses deductible d in coverage j conditional on the observables *and* conditional on $\left(\xi_{r,i}, \xi_{\Omega,i}\right)$ is

$$\begin{aligned} \mathcal{P}(D_{ij}|\xi_{r,i}, \xi_{\Omega,i}) &\equiv \Pr\left(D_{ij} = d | P_{ij}, \widehat{\mu}_{ij}, Z_i, \xi_{r,i}, \xi_{\Omega,i}\right) \qquad \text{(A.8)} \\ &= \Pr\left(\varepsilon_{d'} - \varepsilon_d < U(L_d) - U(L_{d'}) \text{ for all } d' \neq d | P_{ij}, \widehat{\mu}_{ij}, Z_i, \xi_{r,i}, \xi_{\Omega,i}\right) \\ &= \frac{\exp\left(U(L_d)/\sigma\right)}{\sum_{d' \in \mathcal{D}} \exp\left(U(L_{d'})/\sigma\right)}. \end{aligned}$$

The unconditional probability that household i chooses a triple (d_L, d_M, d_H) is given by the integral of the product of equation (A.8) across contexts, against the distribution of $(\xi_{r,i}, \xi_{\Omega,i})$:

$$\int \prod_{k=L,M,H} \mathcal{P}(D_{ik} = d_k | \xi_{r,i}, \xi_{\Omega,i}) \phi(\xi_{r,i}, \xi_{\Omega,i}; 0, \Phi) d\xi_{r,i} d\xi_{\Omega,i}, \tag{A.9}$$

where $\phi(\xi_{r,i}, \xi_{\Omega,i}; 0, \Phi)$ is the bivariate normal density. We observe data $\{D_{ij}, \Gamma_{ij}\}$, where D_{ij} is household i's deductible choice for coverage j and $\Gamma_{ij} \equiv (Z_i, \widehat{\mu}_{ij}, P_{ij})$. In Γ_{ij}, Z_i is a vector of household characteristics, $\widehat{\mu}_{ij}$ is household i's predicted claim probability for coverage j, and P_{ij} denotes household i's menu of premium-deductible pairs for coverage j. The set of fixed parameters (fixed coefficients) to be estimated is $\theta \equiv (\beta_r, \beta_{\Omega,1}, \beta_{\Omega,2}, \beta_{\Omega,3}, \sigma_L, \sigma_M, \sigma_H)$. Additionally, we estimate Φ, the covariance matrix of the heterogeneity terms (random coefficients) $\vec{\xi} \equiv (\xi_{r,i}, \xi_{\Omega,i})$. For a given claim probability μ, equation (A.7) can be seen as a linear function in the explanatory variables $\{p_d, d, (p_d)^2, [(p_d + d)^2 - (p_d)^2]\}$ and their corresponding coefficients, with the coefficient on the last variable constrained to be the product of the second and third coefficients. The first coefficient being one is a scale normalization, since the variance of the choice noise is unconstrained. Hence, identification follows from standard arguments for mixed logit models and our discussion in Section 3.3, provided there is sufficient variation in p and d.[3]

Notice that maximum likelihood estimation cannot be directly carried out, because the integral in equation (A.9) cannot be calculated analytically. Hence, we employ Markov Chain Monte Carlo (MCMC) methods, which avoid integration entirely. Intuitively, in MCMC the integration over unobserved heterogeneity terms is substituted by augmenting the data with draws of the unobserved heterogeneity terms, and updating the distribution from which the draws are taken based on likelihood improvement criteria. The result of the MCMC procedure is the joint posterior distribution of (θ, Φ). By the Bernstein-von Mises theorem, the mean of the posterior is an estimator that, in frequentist terms, is asymptotically equivalent to the (computationally infeasible) maximum likelihood estimator (Train 2009, ch. 12).

The MCMC procedure that we use is built on the Gibbs sampler and Metropolis-Hastings algorithm. It requires the choice of priors, which then are combined with the data, and in particular the observed choices, to obtain the posterior distribution of (θ, Φ). We set:

[3] Furthermore, when we specify $\Omega(\mu) = a + b\mu \exp(\xi_b)$ in Section 6.1, one can immediately see that the model continues to be linear in the explanatory variables $\{p_d, d, (p_d)^2, [(p_d+d)^2 - (p_d)^2], \mu[(p_d+d)^2 - (p_d)^2]\}$ and their corresponding coefficients, with some constraints between the coefficients. Hence, for this case identification also follows from standard arguments for mixed logit models and our discussion in Section 3.3, provided there is sufficient variation in p and d and μ.

Priors

1. The fixed coefficients β_r, $\beta_{\Omega,1}$, $\beta_{\Omega,2}$, and $\beta_{\Omega,3}$ are assumed to have a normal distribution with a diffused prior. The fixed coefficients σ_L, σ_M, and σ_H are assumed to have a lognormal distribution with a diffused prior.

2. The prior on the matrix Φ is inverted Wishart with 2 degrees of freedom and scale matrix $I(2)$.

Initial Values

1. All unobserved heterogeneity terms (random coefficients) are set to zero.

2. The initial value of θ is set to that estimated under the assumption of no unobserved heterogeneity.[4]

3. The initial draw of Φ is $2 \cdot I(2)$.

The posterior distribution of (θ, Φ) conditional on Γ_{ij} is obtained via simulations, using Gibbs sampling and Metropolis-Hastings algorithm, as detailed below.

Simulation: Gibbs Sampling

1. Draws of $\Phi \mid \vec{\xi}_i, \forall i$: The posterior for Φ is inverted Wishart with $2 + N$ degrees of freedom and scale matrix $(2 \cdot I(2) + N \cdot S)/(2 + N)$, where N is the sample size and S is the sample variance of $\vec{\xi}$. A draw from inverted Wishart is obtained according to standard techniques (Train 2009, § 12.5.2).

2. Draws of $\vec{\xi}_i, \forall i \mid \theta, \Phi$: The posterior for each household's $\vec{\xi}$, conditional on θ, on its choices D, and on other observables Γ is

$$K(\vec{\xi}|\theta, \Phi, D, \Gamma) \propto L(D|\vec{\xi}, \theta, \Gamma) \cdot \phi(\vec{\xi}, 0, \Phi),$$

   where

$$L(D|\vec{\xi}, \theta, \Gamma) = \prod_{k=L,M,H} \mathcal{P}(D_{ik} = d_k|\xi_{r,i}, \xi_{\Omega,i})$$

   is the product of the three choice probabilities (across coverages) and $\phi(\vec{\xi}, 0, \Phi)$ is the bivariate normal density. Draws from this posterior are obtained with one step of the Metropolis-Hastings algorithm described below.

[4] For Model 3, the simulation was also run with different starting values for θ, including the one that mimics the case when the model is estimated with no probability distortions.

3. Draws of $\theta \mid \vec{\xi}_i, \forall i$: The posterior for the fixed coefficients θ, conditional on the draws of the random coefficients, $\vec{\xi}_i, \forall i$, on choices D, and on other observables Γ is

$$K(\theta|, \vec{\xi}_i, \forall i, \Phi, D, \Gamma) \propto \prod_i L(D|\vec{\xi}_i, \theta, \Gamma).$$

Draws from this posterior are obtained with one step of the Metropolis Hastings algorithm on the pooled data, as described below.

Metropolis-Hastings Algorithm for Step 2

For each household i, and a given an initial draw $\vec{\xi}_i^0$:

- Draw a bivariate standard normal vector $\vec{\eta}$.
- Create a trial vector $\vec{\xi}_i^1 = \vec{\xi}_i^0 + \rho L \vec{\eta}$, where ρ is a positive scalar and L is the (lower triangular) Choleski factor of Φ.
- Draw a standard uniform variable κ.
- Calculate the ratio

$$F = \frac{L(D|\vec{\xi}^1, \theta, \Gamma) \cdot \phi(\vec{\xi}^1, 0, \Phi)}{L(D|\vec{\xi}^0, \theta, \Gamma) \cdot \phi(\vec{\xi}^0, 0, \Phi)}.$$

- If $\kappa \leq F$, accept the new value of $\vec{\xi}_i^1$. Otherwise, reset $\vec{\xi}_i^1 = \vec{\xi}_i^0$.
- Repeat.

Metropolis-Hastings Algorithm for Step 3

For a given initial draw θ^0:

- Draw a standard normal vector $\vec{\eta}$ of the same dimensionality as θ.
- Create a trial vector $\theta^1 = \theta^0 + \delta \vec{\eta}$, where δ is a positive scalar.
- Draw a standard uniform variable κ.
- Calculate the ratio

$$F = \frac{\prod_i L(D|\vec{\xi}_i, \theta^1, \Gamma)}{\prod_i L(D|\vec{\xi}_i, \theta^0, \Gamma)}.$$

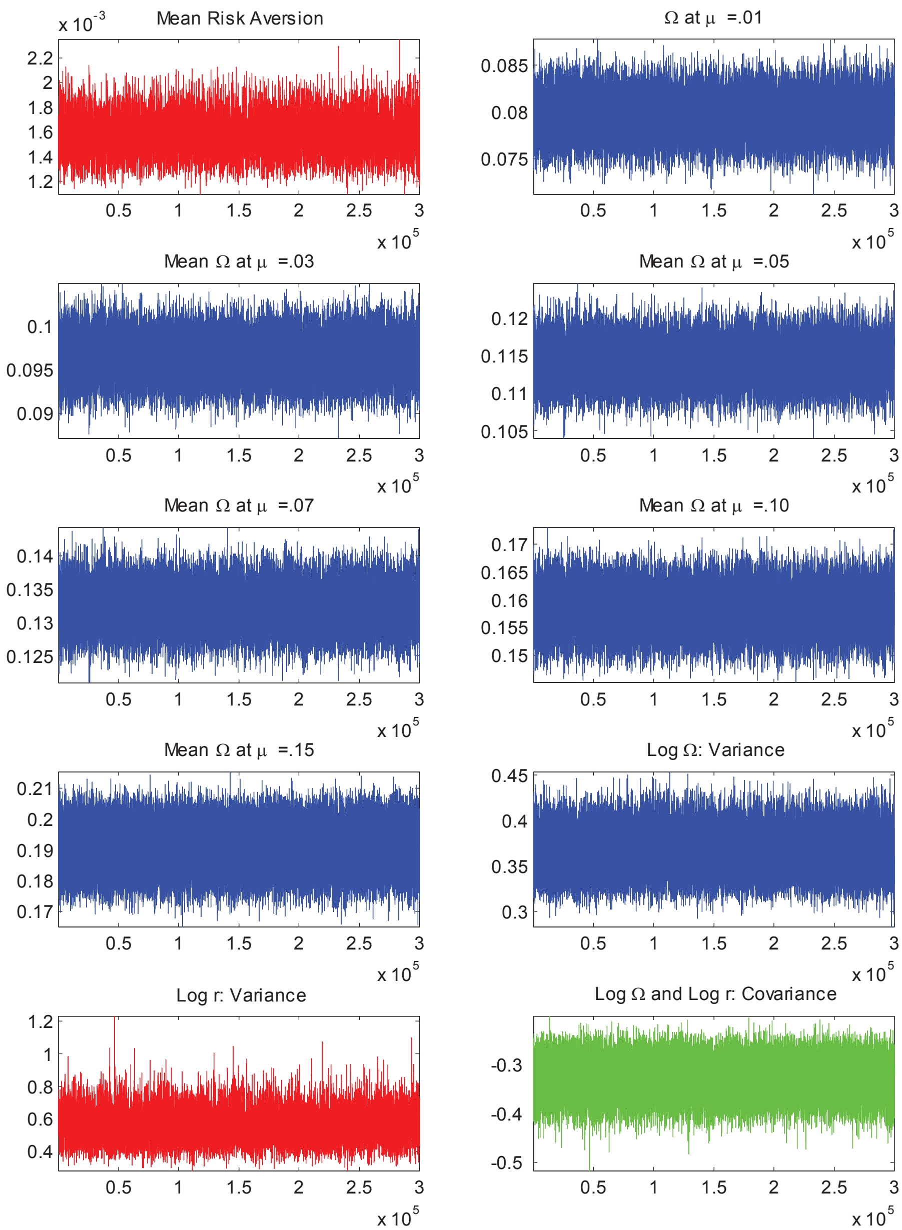
- If $\kappa \leq F$, accept new value of θ^1. Otherwise, reset $\theta^1 = \theta^0$.

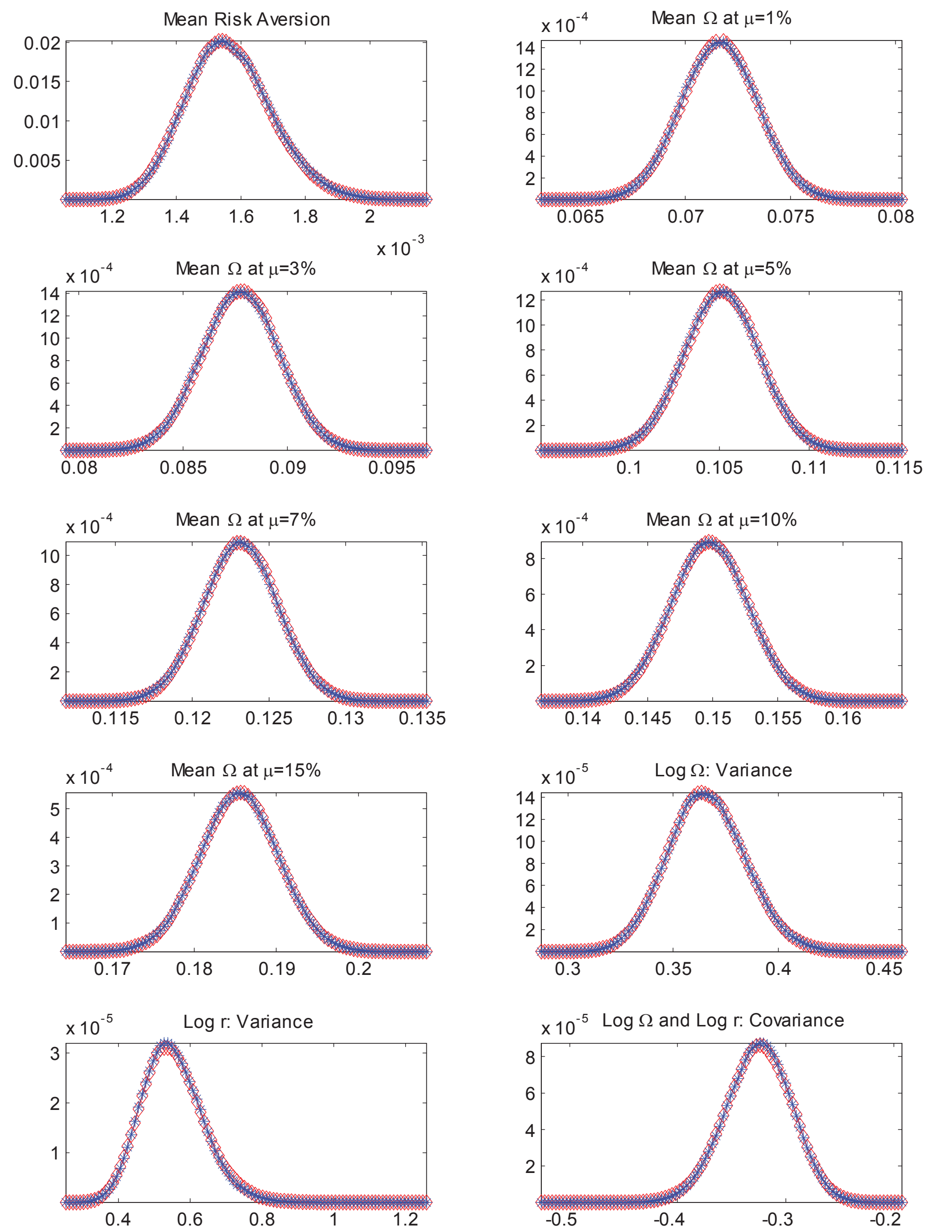
- Repeat.

The Run and Convergence

When running an MCMC procedure, one has to worry that the chain has run sufficiently long to achieve convergence. For Model 3 we run the chain 3,100,000 times.[5] We drop the first 100,000 draws as "burn-in" and retain only every 10th draw as "thinning." The first figure below shows the trace plot for selected objects of interest. The second figure compares the densities of the same objects for the first half of the chain versus the second half.

[5] For Model 4, we run the chain 10,100,000 times.


Mean Risk Aversion
x 10^-3
Ω at μ =.01
Mean Ω at μ =.03
Mean Ω at μ =.05
Mean Ω at μ =.07
Mean Ω at μ =.10
Mean Ω at μ =.15
Log Ω: Variance
Log r: Variance
Log Ω and Log r: Covariance
x 10^5


Mean Risk Aversion
Mean Ω at μ=1%
Mean Ω at μ=3%
Mean Ω at μ=5%
Mean Ω at μ=7%
Mean Ω at μ=10%
Mean Ω at μ=15%
Log Ω: Variance
Log r: Variance
Log Ω and Log r: Covariance

In order to more formally assess convergence of our chains, we employ a battery of statistical tests.[6] We note that in our simulation, as is commonly the case with Metropolis-Hastings based simulations, the draws are autocorrelated. To this end, we use the Raftery-Lewis diagnostics on each chain (i.e., for all parameters) to determine the "burn-in" and "thinning."[7] We then use the second half of the chain, further thinned by a factor of ten, that conforms with the Raftery-Lewis diagnostics and passes the Geweke's chi-squared test under the assumption of iid draws in the chain for each variable.[8]

Core Sample (4170 Households)

	Geweke test (iid)			**Autocorrelations**			
	Mean	**NSE**	**χ^2 pr**	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**
Log r	-6.74	0.00	0.21	0.83	0.47	0.24	-0.01
Log $\Omega(\mu)$: constant	-2.82	0.00	0.24	0.59	0.23	0.10	-0.01
Log $\Omega(\mu)$: linear	10.72	0.01	0.52	0.75	0.25	0.06	-0.04
Log $\Omega(\mu)$: quadratic	-28.07	0.03	0.32	0.68	0.24	0.05	-0.03
σ_L	17.14	0.01	0.65	0.32	0.15	0.05	-0.01
σ_M	10.16	0.00	0.38	0.21	0.08	0.03	0.02
σ_H	95.64	0.07	0.44	0.67	0.38	0.19	-0.01
Φ_r	0.55	0.00	0.77	0.72	0.33	0.17	-0.01
Φ_Ω	0.37	0.00	0.11	0.25	0.13	0.06	0.00
$\Phi_{r,\Omega}$	-0.33	0.00	0.49	0.36	0.09	0.05	-0.01

Raftery-Lewis diagnostics for each parameter chain: I-stat = 3.86.

[6] We use a version of the CODA package adapted for MATLAB by James P. LeSage.

[7] The Raftery-Lewis diagnostic is a run length control diagnostic based on a criterion of accuracy of estimation of the quantile $q = 0.025$. The number of iterations required to estimate the quantile q to within an accuracy of $\pm\ r = 0.01$ with probability $s = 0.95$ is calculated. Separate calculations are performed for each variable within each chain.

[8] The Geweke test is based on the idea that if the sample of draws has attained an equilibrium state, the means of the first 20 percent of the sample of draws versus the last 50 percent of the sample should be roughly the same.

E Additional Sensitivity Checks

In this section, we report the results of several additional sensitivity checks.

E.1 CARA Utility

In our analysis, we consider a second-order Taylor expansion of the utility function, and also CRRA utility. Here we take yet another approach: we assume constant absolute risk aversion (CARA) utility, $u(w) = -\exp(-rw)$. That is, we specify utility as

$$U(L_d) = \frac{EU(L_d)}{u'(w)} = (1 - \Omega(\mu))\,\frac{-\exp(rp_d)}{r} + \Omega(\mu)\frac{-\exp(r(p_d + d))}{r},$$

which we note is independent of wealth w. When we estimate Model 2 with CARA utility, the main message is the same. The estimates for $\Omega(\mu)$ indicate similar probability distortions, albeit somewhat less pronounced than the benchmark, while the estimates for r are higher than the benchmark. See Table A.17.

E.2 Alternative Samples

In the core sample, we restrict attention to households who hold both auto and home policies and who first purchased their auto and home policies from the company in the same year, in either 2005 or 2006. Here we estimate Model 2 using two less restrictive samples: (1) households who hold auto policies and who first purchased their auto policies from the company in the same year, in either 2005 or 2006; and (2) households who hold both auto and home policies and who first purchased their auto and home policies from the company in the same year, in either 2004, 2005, or 2006. Again, the main message is the same. In both samples, the estimates for $\Omega(\mu)$ indicate probability distortions that are very similar to the benchmark. As for standard risk aversion, in sample 1 the estimates for r are higher than the benchmark estimates, while in sample 2 they are lower than the benchmark estimates. See Tables A.18 and A.19.

E.3 Restricted Menus

In our main analysis, we use the full menu of deductible options for each coverage, up to \$1000. In each coverage, however, the vast majority of households choose one of three deductibles: 92.3 percent of households choose a deductible of \$200, \$250, or \$500 in auto collision; 87.1 percent of households choose a deductible of \$200, \$250, or \$500 in auto comprehensive; and 97.5 percent of households choose a deductible of \$250, \$500, or \$1000 in home. Given these choice patterns, one might worry that households do not really consider the other deductible options, which could bias our estimates.[9] To address this concern, we estimate Model 2 when we restrict the menu of deductible options to $\{\$200, \$250, \$500\}$ for each auto coverage and to $\{\$250, \$500, \$1000\}$ for home coverage.[10] The estimates for $\Omega(\mu)$ indicate probability distortions that are similar to the benchmark. Indeed, the overweighting is more pronounced at high claim probabilities. The estimates for r are lower than the benchmark estimates. See Table A.20.

E.4 Alternative Error Structures

In our main analysis, we assume that the utility from every deductible $d \in \mathcal{D}$ is given by $\mathcal{U}(d) = U(L_d) + \varepsilon_d$, where ε_d is an iid Gumbel random variable. Here, we estimate Model 2 under two alternative assumptions (and, for computational and theoretical reasons, using the restricted menus from Section E.3 above): (A) we assume (as before) that the utility from every deductible $d \in \mathcal{D}$ is given by $\mathcal{U}(d) = U(L_d) + \varepsilon_d$, but we assume that ε_d is an iid normal random variable; and (B) we assume that the utility from the maximum deductible, D, is given by $\mathcal{U}(D) = U(L_D) + \varepsilon_D$, where ε_D is an iid normal random variable, but that the utility from the other deductibles are given by $\mathcal{U}(d) = U(L_d) + \zeta_d$, where $\zeta_d = -\varepsilon_D$ for the minimum deductible and $\zeta_d = 0$ for the intermediate deductible. Alternative A provides a check of the Gumbel error assumption. Alternative B adds a check of the iid assumption. More specifically, we consider alternative B to address concerns arising from the fact that in principle the iid assumption allows for nonmonotonic ranking of deductibles. Once again, the main message is the same. Under both alternatives, the estimates for $\Omega(\mu)$ indicate similar probability distortions, though generally somewhat more pronounced. In addition, under alternative A the estimates for r are lower than the benchmark estimates, while under alternative B they are higher than the benchmark estimates. See Tables A.21 and A.22.

F Appendix Tables and Figures

On the ensuing pages, we report Tables A.1 through A.22 and Figures A.1 through A.4.

[9] For instance, when a household chooses a \$250 deductible in home, we are using the fact that it did not choose a \$100 deductible to infer an upper bound on its aversion to risk. But if the household in fact does not even consider the \$100 deductible as an option, our inference would be invalid. Similarly, when a household chooses a \$500 deductible in auto comprehensive, we are using the fact that it did not choose a \$1000 deductible to infer a lower bound on its aversion to risk. Again, if the household in fact does not even consider the \$1000 deductible as an option, our inference would be invalid.

[10] In each case, if a household's actual deductible choice is outside the restricted menu, we assign to the household the deductible option from the restricted menu that is closest to their actual deductible choice. In this respect, we follow Cohen and Einav (2007).

Table A.1: Summary of Premium Menus - Auto Collision

Core Sample (4170 Households)

	Deductible Choice					
	$100	**$200**	**$250**	**$500**	**$1000**	**All**
Mean annual premium for coverage with $500 deductible	110	129	146	189	255	180
Standard deviation	54	54	66	96	168	100
Mean cost of decreasing deductible from $500 to $250	33	38	44	57	77	54
Standard deviation	17	17	20	29	52	31
Mean savings from increasing deductible from $500 to $1000	24	29	33	43	58	41
Standard deviation	12	12	15	22	39	23
Number of households	42	559	467	2822	280	4170

Note: All values in dollars, except number of households.

Table A.2: Summary of Premium Menus - Auto Comprehensive

Core Sample (4170 Households)

	Deductible Choice						
	$50	**$100**	**$200**	**$250**	**$500**	**$1000**	**All**
Mean annual premium for coverage with $500 deductible	61	70	92	98	136	258	115
Standard deviation	27	33	43	41	71	247	81
Mean cost of decreasing deductible from $500 to $250	16	18	24	26	36	68	30
Standard deviation	7	9	11	11	19	66	22
Mean savings from increasing deductible from $500 to $1000	12	14	18	19	27	51	23
Standard deviation	5	7	9	8	14	49	16
Number of households	216	171	1397	440	1795	151	4170

Note: All values in dollars, except number of households.

Table A.3: Summary of Premium Menus - Home

Core Sample (4170 Households)

	Deductible Choice						
	$100	**$250**	**$500**	**$1000**	**$2500**	**$5000**	**All**
Mean annual premium for coverage with $500 deductible	366	520	631	972	2218	3366	679
Standard deviation	113	218	308	593	2289	1808	519
Mean cost of decreasing deductible from $500 to $250	31	42	52	80	183	275	56
Standard deviation	6	18	26	48	201	140	43
Mean savings from increasing deductible from $500 to $1000	41	57	69	107	244	368	74
Standard deviation	8	23	34	64	268	188	58
Number of households	36	1239	2166	664	50	15	4170

Note: All values in dollars, except number of households.

Table A.4: Claim Rate Regressions - Auto
Poisson Panel Regression Model with Random Effects
Full Data Set (1,348,020 Household-Year Records)

	Collision		Comprehensive	
	Coef	**Std Err**	**Coef**	**Std Err**
Constant	-6.7646 **	0.0616	-7.9277 **	0.1057
Driver 2 Indicator	-0.0485	0.0593	-0.3542 **	0.1022
Driver 3+ Indicator	0.3215 **	0.0733	-0.1261	0.1201
Vehicle 2 Indicator	0.5991 **	0.0466	0.6502 **	0.0782
Vehicle 3+ Indicator	0.7312 **	0.0596	0.8766 **	0.0937
Young Driver	-0.0058	0.0296	0.0895 **	0.0453
Driver 1 Age	-0.0210 **	0.0015	0.0113 **	0.0029
Driver 1 Age Squared	0.0002 **	0.0000	-0.0002 **	0.0000
Driver 1 Female	0.1040 **	0.0093	-0.0672 **	0.0168
Driver 1 Married	0.0630 **	0.0111	0.0640 **	0.0201
Driver 1 Divorced	0.0186	0.0141	0.0914 **	0.0247
Driver 1 Separated	0.0392	0.0256	0.0791	0.0428
Driver 1 Single	.	.	.	.
Driver 1 Widowed	0.0031	0.0160	-0.0170	0.0335
Vehicle 1 Age	-0.0354 **	0.0019	-0.0286 **	0.0030
Vehicle 1 Age Squared	-0.0006 **	0.0001	0.0000	0.0002
Vehicle 1 Business	.	.	.	.
Vehicle 1 Farm	-0.2575 **	0.0872	0.0206	0.1194
Vehicle 1 Pleasure	-0.1094 **	0.0306	-0.1118 **	0.0526
Vehicle 1 Work	-0.0831 **	0.0304	-0.0620	0.0523
Vehicle 1 Passive Restraint	-0.1087 **	0.0239	-0.0858 **	0.0352
Vehicle 1 Anti-Theft	0.0754 **	0.0078	0.0735 **	0.0136
Vehicle 1 Anti-Lock	0.0581 **	0.0080	0.0729 **	0.0139
Driver 2 Age	0.0115 **	0.0024	0.0181 **	0.0042
Driver 2 Age Squared	-0.0001 **	0.0000	-0.0001 **	0.0000
Driver 2 Female	0.1204 **	0.0151	-0.0376	0.0257
Driver 2 Married	-0.0835 **	0.0191	-0.0408	0.0326
Driver 2 Divorced	-0.1579	0.1027	-0.1347	0.1636
Driver 2 Separated	0.0254	0.2130	0.1796	0.3226
Driver 2 Single	.	.	.	.
Driver 2 Widowed	-0.0802	0.1383	-1.1835 **	0.3864
Vehicle 2 Age	-0.0332 **	0.0016	-0.0229 **	0.0027
Vehicle 2 Age Squared	0.0004 **	0.0001	0.0002 **	0.0001
Vehicle 2 Business	.	.	.	.
Vehicle 2 Farm	-0.1703	0.1056	-0.1345	0.1500
Vehicle 2 Pleasure	-0.1805 **	0.0380	-0.0563	0.0663
Vehicle 2 Work	-0.1670 **	0.0381	0.0119	0.0664
Vehicle 2 Passive Restraint	-0.0428 **	0.0201	-0.0875 **	0.0294
Vehicle 2 Anti-Theft	0.0547 **	0.0103	0.0385 **	0.0171
Vehicle 2 Anti-Lock	0.0317 **	0.0105	0.0199	0.0170
Insurance Score	-0.0017 **	0.0000	-0.0013 **	0.0001
Previous Accident	0.0913 **	0.0156	0.0756 **	0.0277
Previous Convictions	0.1476	0.0888	0.0648	0.1670
Previous Reinstated	0.0170	0.0558	0.0003	0.0996
Previous Revocation	-0.0218	0.1456	0.3156	0.1967
Previous Suspension	0.0463	0.0564	0.0125	0.1026
Previous Violation	0.0827 **	0.0093	0.0577 **	0.0161
Year Dummies	Yes		Yes	
Territory Codes	Yes		Yes	
Variance (ϕ)	0.2242 **	0.0065	0.5661 **	0.0198
Loglikelihood	-399,318		-169,817	

Note: Territory codes indicate rating territories, which are based on actuarial risk factors such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services.

** Significant at 5 percent level.

Table A.5: Claim Rate Regression - Home

Poisson Panel Regression Model with Random Effects

Full Data Set (1,265,229 Household-Year Records)

	Coef		Std Err
Constant	-7.3642	**	0.0978
Dwelling Value	0.0000	**	0.0000
Home Age	0.0016	**	0.0006
Home Age Squared	0.0000	**	0.0000
Number of Families	-0.0021		0.0023
Distance to Hydrant	0.0000		0.0000
Alarm Discount	0.2463	**	0.0195
Protection Devices	-0.1852	**	0.0239
Farm/Business	0.1044	**	0.0242
Primary Home	0.4832	**	0.0819
Owner Occupied	0.2674	**	0.0419
Construction: Fire Resistant	0.1525		0.1342
Construction: Masonry	0.0751	**	0.0172
Construction: Masonry/Veneer	0.0755	**	0.0252
Construction: Frame	.		.
Insurance Score	-0.0026	**	0.0000
Year Dummies		Yes	
Protection Classes		Yes	
Territory Codes		Yes	
Variance (ϕ)	0.4514	**	0.0086
Loglikelihood		-347,278	

Notes: Territory codes indicate rating territories, which are based on actuarial risk factors such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services. Protection classes gauge the effectiveness of local fire protection and building codes.

** Significant at 5 percent level.

Table A.6: Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-7.39 **	0.09	-2.73 **	0.02	12.40 **	0.41	-35.61 **	2.47
Driver 1 age	-1.47 **	0.14	0.18 **	0.05	2.52 **	0.48	1.00 **	0.31
Driver 1 age squared	1.09 **	0.15	0.00	0.05	-4.94 **	0.43	9.92 **	1.82
Driver 1 female	0.15 **	0.04	-0.05 **	0.01	1.57 **	0.28	-12.29 **	1.66
Driver 1 single	0.08	0.05	-0.01	0.01	0.77 **	0.26	-5.93 **	1.91
Driver 1 married	0.09	0.06	-0.03	0.02	1.40 **	0.27	-9.34 **	1.40
Insurance score	-0.15 **	0.05	-0.02	0.01	2.31 **	0.21	-11.00 **	1.23
Driver 2 indicator	-0.04	0.06	0.00	0.02	-1.44 **	0.38	7.63 **	2.19
Mean fitted value	0.00073		-2.73		12.40		-35.61	
Median fitted value	0.00056		-2.73		12.42		-34.46	
σ_L	27.22 **	0.76						
σ_M	17.91 **	0.48						
σ_H	65.45 **	2.68						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.7: Model 3

Core Sample (4170 Households)

	Estimate		Std Err
r	0.00118	***	0.00011
Log Ω(μ): constant	-2.82	***	0.03
Log Ω(μ): linear	10.72	***	0.52
Log Ω(μ): quadratic	-28.04	***	3.46
σ_L	17.14	***	0.61
σ_M	10.16	***	0.38
σ_H	95.69	***	6.74
Φ_r	0.55	***	0.09
Φ_Ω	0.37	***	0.02
$\Phi_{r,\Omega}$	-0.33	***	0.03
Mean fitted r	0.00156		
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)	-0.72		

*** Significant at 1 percent level.

Table A.8: Model 4

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-6.81 **	0.10	-2.83 **	0.03	11.22 **	0.50	-31.67 **	3.42
Driver 1 age	-0.41 **	0.06	0.18 **	0.02	-2.01 **	0.51	7.14 **	3.35
Driver 1 age squared	0.07	0.04	0.01	0.02	-0.66	0.44	1.96	2.80
Driver 1 female	0.06	0.05	-0.05	0.03	2.03 **	0.59	-14.48 **	3.77
Driver 1 single	-0.02	0.05	0.02	0.03	0.24	0.60	-3.65	3.87
Driver 1 married	-0.03	0.08	0.00	0.04	1.06	0.81	-6.68	5.07
Insurance score	-0.14 **	0.05	0.00	0.02	1.27 **	0.55	-5.43	3.43
Driver 2 indicator	-0.13	0.07	0.04	0.03	-1.84 **	0.72	7.59	4.59
Mean fitted value	0.00147		-2.83		11.22		-31.67	
Median fitted value	0.00145		-2.83		11.04		-28.53	
σ_L	17.09 **	0.61						
σ_M	10.46 **	0.40						
σ_H	90.34 **	6.23						
Φ_r	0.58 **	0.09						
Φ_Ω	0.35 **	0.02						
$\Phi_{r,\Omega}$	-0.33 **	0.03						
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)	-0.72							

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.9: Unobserved Heterogeneity in Risk – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-7.11 **	0.08	-2.73 **	0.02	11.04 **	0.41	-21.78 **	1.95
Driver 1 age	-1.22 **	0.19	0.15 **	0.07	3.31 **	0.57	-3.86 **	1.66
Driver 1 age squared	0.75 **	0.20	0.02	0.06	-4.92 **	0.61	7.64 **	1.10
Driver 1 female	0.05	0.05	-0.01	0.02	1.24 **	0.29	-9.90 **	1.32
Driver 1 single	0.03	0.07	0.00	0.02	0.56 **	0.21	-3.25 **	1.20
Driver 1 married	0.02	0.12	-0.01	0.05	1.24 **	0.46	-6.87 **	1.00
Insurance score	-0.12 **	0.05	-0.01	0.03	1.49 **	0.62	-3.64	2.73
Driver 2 indicator	-0.09	0.08	0.01	0.04	-1.40	0.75	5.27 **	1.99
Mean fitted value	0.00097		-2.73		11.04		-21.78	
Median fitted value	0.00076		-2.74		11.05		-19.49	
σ_L	25.93 **	0.80						
σ_M	18.02 **	0.56						
σ_H	66.01 **	2.78						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.10: Unobserved Heterogeneity in Risk – Model 3

Core Sample (4170 Households)

	Estimate		Std Err
r	0.00134	***	0.00012
Log $\Omega(\mu)$: constant	-2.77	***	0.09
Log $\Omega(\mu)$: linear	8.76	***	0.03
Log $\Omega(\mu)$: quadratic	-16.70	***	0.08
σ_L	17.00	***	0.04
σ_M	9.72	***	0.61
σ_H	105.49	***	0.38
Φ_r	0.42		7.99
Φ_Ω	0.37	***	0.07
$\Phi_{r,\Omega}$	-0.28	***	0.02
Mean fitted r	0.00166		
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)	-0.72		

*** Significant at 1 percent level.

Table A.11: Unobserved Heterogeneity in Risk
Core Sample (4170 Households)

	Estimate		Std Err
r	0.00160	***	0.00020
Ω(μ): intercept	0.04	***	0.00
Ω(μ): slope	0.86	***	0.03
σ_L	21.83	***	0.97
σ_M	10.98	***	0.50
σ_H	89.33	***	7.01
Ψ_r	0.67	***	0.09
Ψ_L	0.76	***	0.05
Ψ_M	3.11	***	0.16
Ψ_H	0.88	***	0.07
$\Psi_{r,L}$	-0.33	***	0.06
$\Psi_{r,M}$	-0.69	***	0.11
$\Psi_{r,H}$	0.17	***	0.06
$\Psi_{L,H}$	0.40	***	0.05
$\Psi_{M,H}$	0.85	***	0.10
$\Psi_{L,M}$	1.44	***	0.08
Mean fitted r	0.00224		
Implied corr($\xi_{r,i}$,$\xi_{b,Li}$)	-0.45		
Implied corr($\xi_{r,i}$,$\xi_{b,Mi}$)	-0.47		
Implied corr($\xi_{r,i}$,$\xi_{b,Hi}$)	0.22		
Implied corr($\xi_{b,Li}$,$\xi_{b,Mi}$)	0.94		
Implied corr($\xi_{b,Li}$,$\xi_{b,Hi}$)	0.49		
Implied corr($\xi_{b,Mi}$,$\xi_{b,Hi}$)	0.51		

*** Significant at 1 percent level.

Table A.12: Restricted Choice Noise – Model 2
Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-14.37 **	0.14	-2.52 **	0.00	12.07 **	0.07	-36.63 **	2.08
Driver 1 age	1.89 **	0.01	0.18 **	0.00	3.11 **	0.01	-1.65 **	0.03
Driver 1 age squared	-6.73 **	0.07	0.00	0.00	-4.91 **	0.02	11.08 **	1.07
Driver 1 female	0.00	0.01	-0.05 **	0.00	1.24 **	0.01	-9.66 **	0.78
Driver 1 single	0.20 **	0.00	-0.01 **	0.00	0.35 **	0.00	-3.36 **	0.03
Driver 1 married	0.21 **	0.00	-0.03 **	0.00	1.05 **	0.00	-7.29 **	0.30
Insurance score	-0.05 **	0.01	-0.02 **	0.00	2.07 **	0.01	-9.91 **	0.54
Driver 2 indicator	-0.27 **	0.01	0.00 **	0.00	-1.49 **	0.02	8.37 **	0.32
Mean fitted value	0.00005		-2.52		12.07		-36.63	
Median fitted value	0.00000		-2.53		12.13		-36.25	
σ_L	13.61							
σ_M	8.95							
σ_H	32.72							

Notes: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*). The variance terms σ_L, σ_M, and σ_H are by construction.

** Significant at 5 percent level.

Table A.13: Unobserved Heterogeneity in Risk – Model 3

Core Sample (4170 Households)

	Estimate		**Std Err**
r	0.00000		0.00000
Log $\Omega(\mu)$: constant	-2.40	***	0.01
Log $\Omega(\mu)$: linear	10.17	***	0.20
Log $\Omega(\mu)$: quadratic	-37.00	***	1.44
σ_L	8.57		
σ_M	5.08		
σ_H	47.84		
Φ_r	2.77		3.47
Φ_Ω	0.25	***	0.01
$\Phi_{r,\Omega}$	0.05		0.26
Mean fitted r	0.00000		
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)	0.07		

Note: The variance terms σ_L, σ_M, and σ_H are by construction.

*** Significant at 1 percent level.

Table A.14: CRRA Utility – Model 2

Core Sample (4170 Households)

	ρ		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-1.16 **	0.15	-2.51 **	0.02	12.31 **	0.46	-39.53 **	3.01
Driver 1 age	-0.44	1.00	-0.10	0.11	1.56	3.08	5.04	20.31
Driver 1 age squared	-0.05	0.99	0.20	0.11	-3.61	3.09	5.90	20.43
Driver 1 female	-0.16	0.15	-0.01	0.02	1.14 **	0.56	-8.48 **	3.66
Driver 1 single	-0.30	0.17	0.02	0.02	0.18	0.55	-1.23	3.65
Driver 1 married	-0.07	0.24	-0.01	0.03	1.01	0.83	-6.40	5.49
Insurance score	-0.11	0.13	-0.03	0.02	1.74 **	0.51	-8.36 **	3.33
Driver 2 indicator	-0.30	0.21	0.01	0.03	-1.32	0.71	7.36	4.71
Mean fitted value	0.37		-2.51		12.31		-39.53	
Median fitted value	0.31		-2.53		12.35		-39.31	
σ_L	22.79 **	0.68						
σ_M	15.77 **	0.52						
σ_H	40.24 **	1.21						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.15: CRRA Utility – Model 3

Core Sample (4170 Households)

	Estimate		Std Err
ρ	0.18	***	0.03
Log $\Omega(\mu)$: constant	-2.47	***	0.02
Log $\Omega(\mu)$: linear	11.39	***	0.47
Log $\Omega(\mu)$: quadratic	-37.86	***	3.23
σ_L	13.21	***	0.38
σ_M	8.09	***	0.27
σ_H	42.41	***	1.37
Φ_r	0.27	***	0.07
Φ_Ω	0.18	***	0.01
$\Phi_{r,\Omega}$	-0.12	***	0.02
Mean fitted ρ	0.21		
Implied corr($\xi_{r,i}$,$\xi_{\Omega,i}$)	-0.56		

*** Significant at 1 percent level.

Table A.16: All Claims $750 – Model 1a

Core Sample (4170 Households)

	Estimate		Std Err
r	0.00195	***	0.00000
Log $\Omega(\mu)$: constant	-2.73	***	0.03
Log $\Omega(\mu)$: linear	12.45	***	0.30
Log $\Omega(\mu)$: quadratic	-32.19	***	2.17
σ_L	54.30	***	2.57
σ_M	28.08	***	0.47
σ_H	244.69	***	0.00

*** Significant at 1 percent level.

Table A.17: CARA Utility – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-6.83 **	0.05	-3.01 **	0.03	11.94 **	0.65	-34.82 **	3.94
Driver 1 age	-0.75 **	0.34	0.33 **	0.15	8.43 **	1.33	-18.69 **	2.73
Driver 1 age squared	0.53	0.39	-0.08	0.16	-10.42 **	1.49	28.08 **	3.37
Driver 1 female	0.04	0.02	-0.02	0.03	0.90	0.75	-9.00 **	4.25
Driver 1 single	0.02	0.02	-0.01	0.02	0.34	0.23	-3.13	1.90
Driver 1 married	-0.02	0.03	0.02	0.04	0.67	0.89	-4.54	2.74
Insurance score	-0.04	0.03	-0.02	0.03	2.30 **	0.78	-10.59 **	3.83
Driver 2 indicator	-0.05	0.03	0.01	0.03	-0.95 **	0.46	6.21 **	1.57
Mean fitted value	0.00113		-3.01		11.94		-34.82	
Median fitted value	0.00103		-3.01		12.06		-34.67	
σ_L	33.31 **	0.98						
σ_M	21.30 **	0.64						
σ_H	104.30 **	5.29						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.18: Alternative Sample 1 – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Constant	-6.83 **	0.04	-2.88 **	0.01	14.01 **	0.34	-43.10 **	2.78
Driver 1 age	1.14 **	0.34	-0.31 **	0.09	-1.14 **	0.54	-7.11 **	2.75
Driver 1 age squared	-1.28 **	0.36	0.41 **	0.08	-2.79 **	0.34	26.03 **	1.11
Driver 1 female	0.16 **	0.02	-0.01	0.01	-0.89 **	0.22	2.40 **	0.86
Driver 1 single	-0.08 **	0.04	0.03 **	0.02	0.14	0.21	-1.65	1.27
Driver 1 married	0.01	0.03	0.03 **	0.01	-0.97 **	0.14	2.68 **	0.32
Insurance score	-0.07 **	0.02	0.03 **	0.01	0.86 **	0.25	-4.51 **	1.21
Driver 2 indicator	0.10 **	0.03	0.01	0.01	-2.23 **	0.30	10.63 **	2.04
Mean fitted value	0.00113		-2.88		14.01		-43.10	
Median fitted value	0.00112		-2.92		14.26		-45.60	
σ_L	31.17 **	0.48						
σ_M	17.91 **	0.43						
σ_H	18.93 **	0.27						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.19: Alternative Sample 2 – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Constant	-7.57 **	0.08	-2.63 **	0.02	12.91 **	0.43	-42.12 **	2.97
Driver 1 age	-1.31 **	0.21	0.06	0.09	4.52	2.46	-11.77	15.32
Driver 1 age squared	0.91 **	0.22	0.08	0.09	-6.73 **	2.53	22.96	16.35
Driver 1 female	0.15 **	0.04	-0.03	0.02	0.94 **	0.29	-7.90 **	1.65
Driver 1 single	0.05	0.05	0.02	0.02	-0.10	0.33	-0.18	2.01
Driver 1 married	-0.13	0.08	0.04 **	0.02	0.42	0.45	-4.30	2.58
Insurance score	-0.08 **	0.04	-0.01	0.02	2.19 **	0.38	-12.41 **	2.31
Driver 2 indicator	0.13	0.07	-0.03	0.02	-1.32 **	0.41	8.32 **	2.86
Mean fitted value	0.00060		-2.63		12.91		-42.12	
Median fitted value	0.00048		-2.65		12.99		-42.40	
σ_L	28.61 **	0.71						
σ_M	18.78 **	0.47						
σ_H	58.74 **	2.00						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.20: Restricted Menus – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**	**Estimate**	**Std Err**
Constant	-8.79 **	0.27	-2.81 **	0.04	12.88 **	0.65	-30.29 **	3.12
Driver 1 age	-4.22 **	0.85	0.03	0.17	6.47 **	2.37	-8.09	8.02
Driver 1 age squared	3.92 **	0.94	0.38 **	0.17	-10.94 **	2.28	23.24 **	7.44
Driver 1 female	0.14	0.16	0.03	0.04	0.91	1.00	-9.94 **	5.00
Driver 1 single	0.18	0.19	-0.01	0.04	0.96	0.99	-6.98	5.07
Driver 1 married	0.20	0.27	-0.02	0.04	1.31	0.76	-8.92 **	3.09
Insurance score	-0.54 **	0.17	-0.04	0.03	2.59 **	0.66	-10.26 **	2.93
Driver 2 indicator	-0.07	0.21	0.00	0.04	-1.19	0.79	5.52	4.05
Mean fitted value	0.00029		-2.81		12.88		-30.29	
Median fitted value	0.00013		-2.88		13.46		-30.35	
σ_L	26.26 **	0.84						
σ_M	19.83 **	0.89						
σ_H	76.56 **	4.49						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.21: Alternative Error Structure A – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Constant	-9.22 **	0.20	-2.81 **	0.05	13.48 **	0.80	-32.52 **	3.46
Driver 1 age	-4.89 **	0.97	0.04	0.20	6.23 **	1.69	-6.01	5.40
Driver 1 age squared	4.61 **	1.22	0.35	0.24	-10.31 **	2.68	19.36 **	2.34
Driver 1 female	0.09	0.10	0.03	0.02	0.91 **	0.27	-9.67 **	2.15
Driver 1 single	0.30 **	0.13	-0.01	0.02	0.84 **	0.37	-5.96 **	1.87
Driver 1 married	0.44 **	0.13	-0.02	0.03	1.22 **	0.32	-8.34 **	1.26
Insurance score	-0.61 **	0.13	-0.04	0.03	2.57 **	0.76	-10.16 **	3.30
Driver 2 indicator	-0.07	0.18	0.00	0.03	-1.18 **	0.47	5.43 **	2.14
Mean fitted value	0.00022		-2.81		13.48		-32.52	
Median fitted value	0.00008		-2.87		13.96		-32.32	
σ_L	51.63 **	1.34						
σ_M	36.49 **	1.46						
σ_H	130.57 **	6.96						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

Table A.22: Alternative Error Structure B – Model 2

Core Sample (4170 Households)

	r		Log Ω(μ): constant		Log Ω(μ): linear		Log Ω(μ): quadratic	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Constant	-7.04 **	0.09	-2.04 **	0.03	3.21 **	0.64	-7.30 **	3.34
Driver 1 age	-1.43 **	0.18	0.42 **	0.11	2.89	1.90	-7.31	5.79
Driver 1 age squared	1.16 **	0.19	-0.28 **	0.12	-3.60	2.27	10.37	8.06
Driver 1 female	0.00	0.03	0.03	0.02	0.08	0.50	-2.96	1.98
Driver 1 single	-0.05	0.04	-0.09 **	0.04	2.03	1.11	-7.82	5.07
Driver 1 married	-0.11 **	0.06	0.01	0.05	1.32	1.01	-7.73 **	3.23
Insurance score	-0.26 **	0.03	-0.01	0.01	2.25 **	0.29	-8.80 **	1.26
Driver 2 indicator	-0.07	0.05	-0.02	0.02	0.14	0.22	2.25	2.11
Mean fitted value	0.00101		-2.04		3.21		-7.30	
Median fitted value	0.00081		-2.01		3.36		-7.54	
σ_L	62.75 **	2.22						
σ_M	53.33 **	3.23						
σ_H	65.80 **	2.94						

Note: Each independent variable *z* is normalized as (*z* -mean(*z*))/stdev(*z*).

** Significant at 5 percent level.

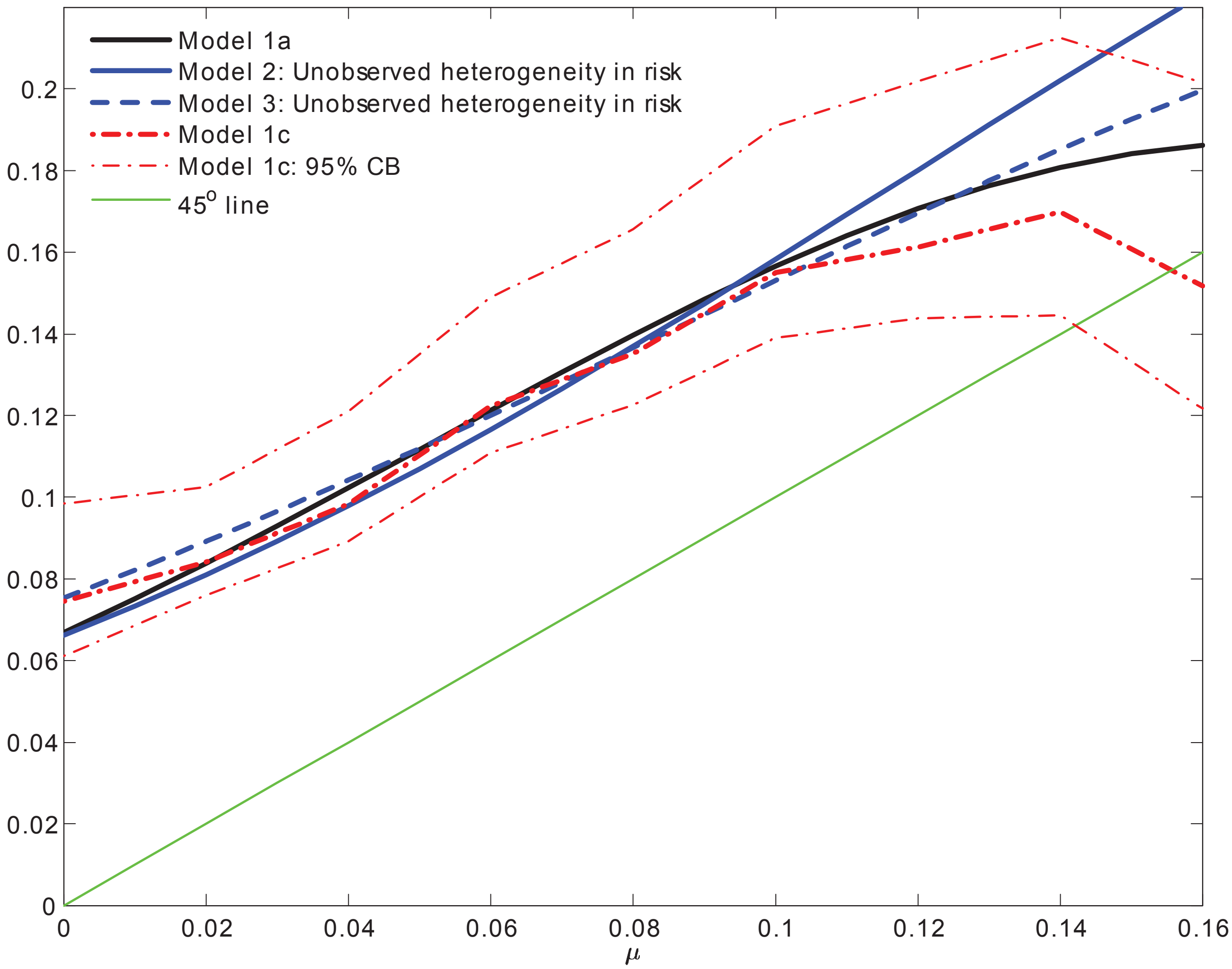


Figure A.1: Mean Estimated $\Omega(\mu)$ – Unobserved Heterogeneity in Risk

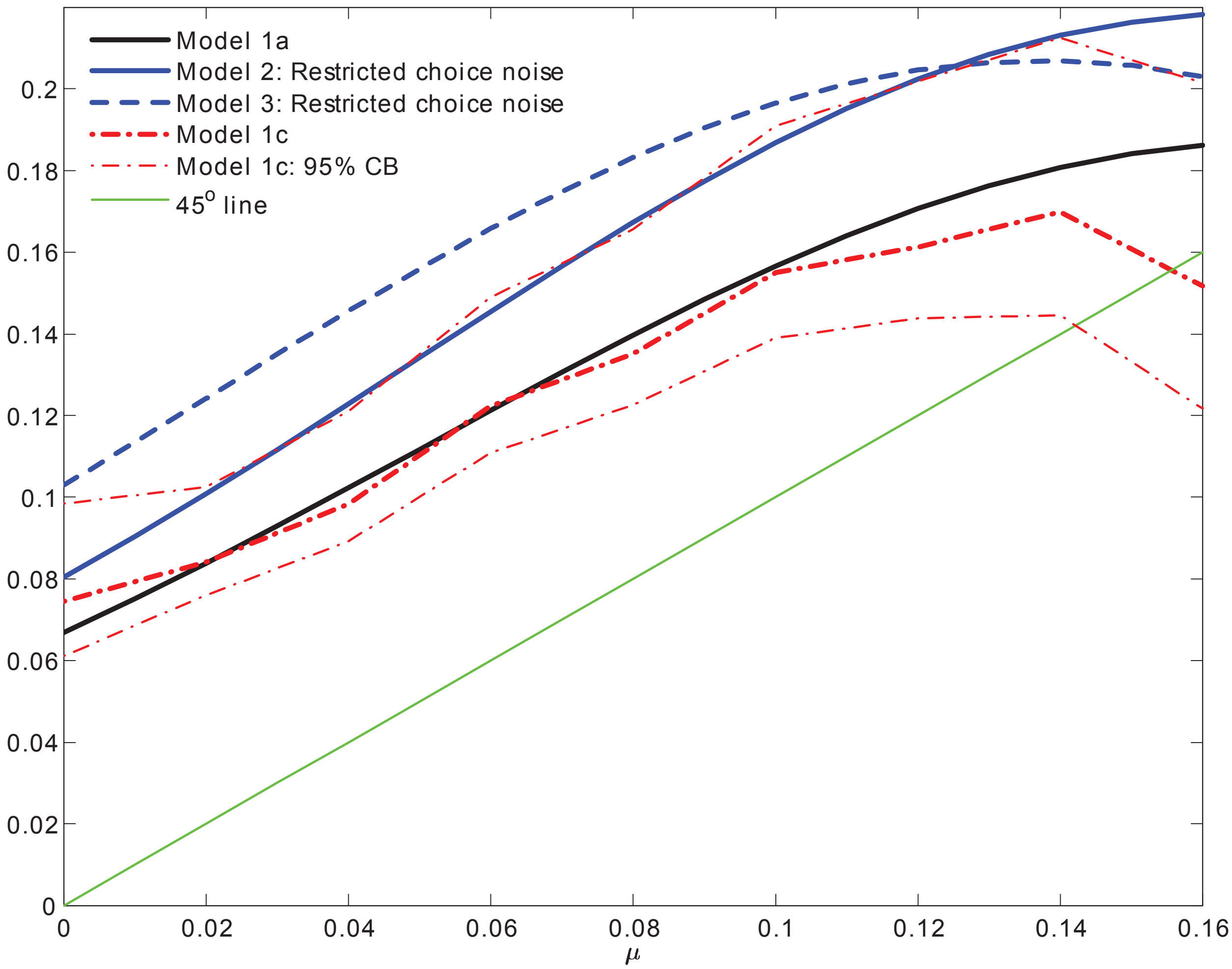


Figure A.2: Mean Estimated $\Omega(\mu)$ – Restricted Choice Noise

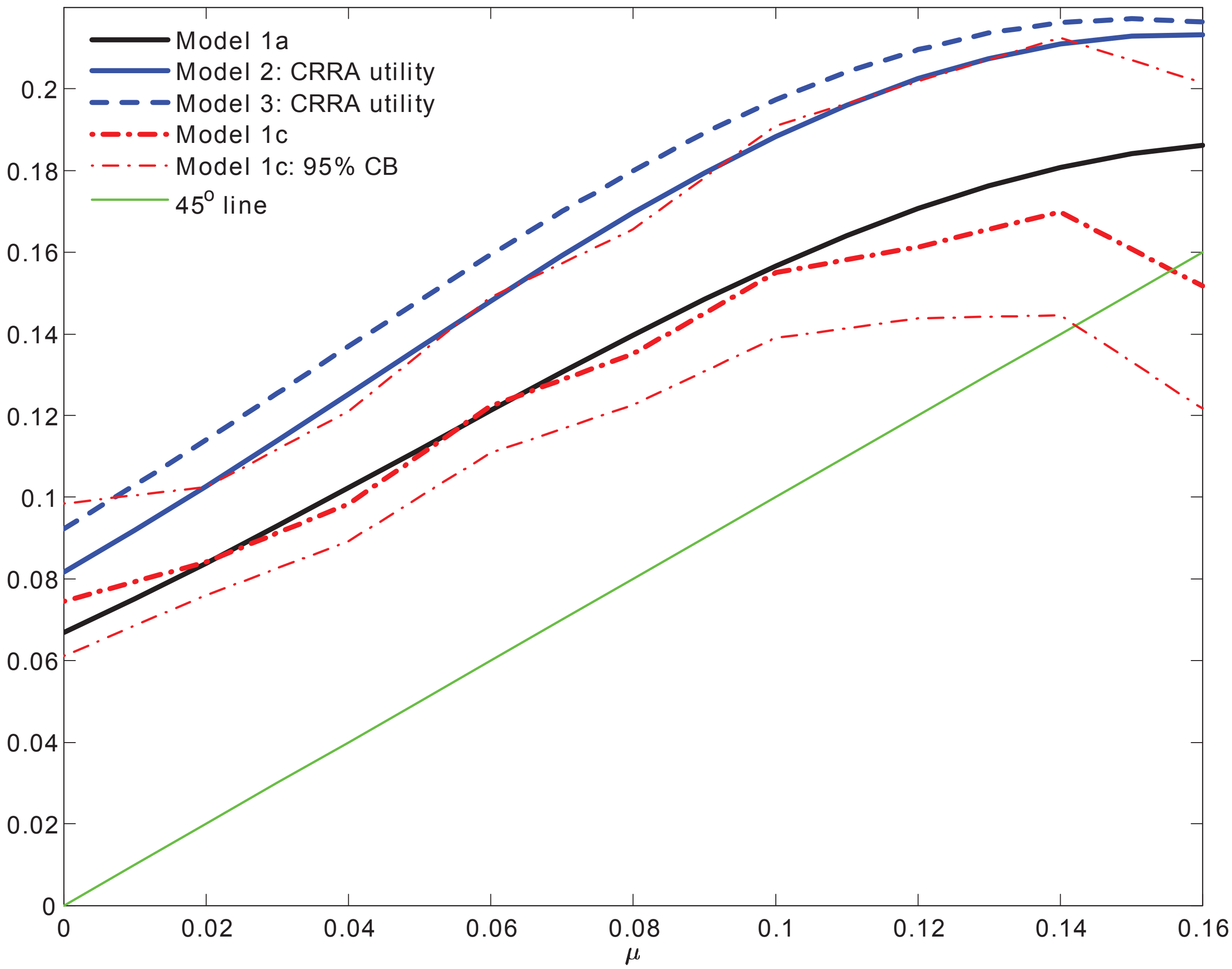


Figure A.3: Mean Estimated $\Omega(\mu)$ – CRRA Utility

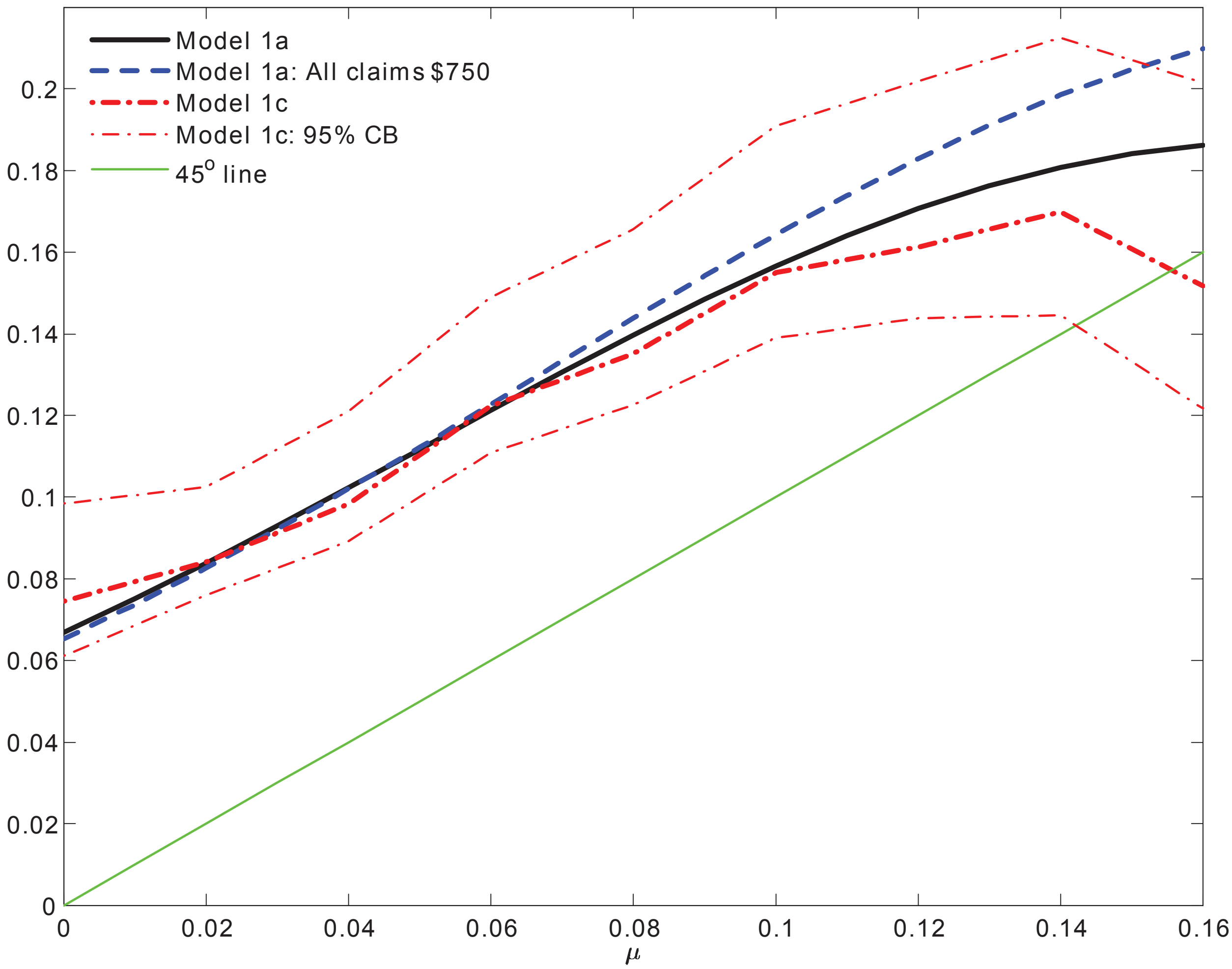


Figure A.4: Estimated $\Omega(\mu)$ – All Claims $750