

Are Risk Preferences Stable across Contexts? Evidence from Insurance Data

By LEVON BARSEGHYAN, JEFFREY PRINCE, AND JOSHUA C. TEITELBAUM*

Using a unique dataset, we test whether households' deductible choices in auto and home insurance reflect stable risk preferences. Our test relies on a structural model that assumes households are objective expected utility maximizers and claims are generated by household-coverage specific Poisson processes. We find that the hypothesis of stable risk preferences is rejected by the data. Our analysis suggests that many households exhibit greater risk aversion in their home deductible choices than their auto deductible choices. Our results are robust to several alternative modeling assumptions. (JEL D11, D83)

Economists generally assume that risk preferences are stable across decision contexts. The assumption of context-invariant risk preferences implies that multiple risky choices by the same economic unit should reflect the same degree of risk aversion (or risk loving), even if the contexts of the decisions are different.¹ This idea underlies standard models in many fields of economics, including finance, industrial organization, insurance, labor economics, and macroeconomics. It also motivates an important literature in empirical microeconomics—estimating risk preferences from data on choices in a single context. What is the empirical validity of this assumption? Are estimates of risk preferences derived from choices in one context relevant for another context?

This paper addresses these questions by examining households' insurance choices. We assemble and use a new dataset that records choices made by a sample of households in connection with three types of insurance coverage: auto collision, auto comprehensive, and home all perils. Among other things, the data contain households' deductible choices for each type of coverage. The data also include

*Barseghyan: Department of Economics, Cornell University, 456 Uris Hall, Ithaca, NY 14853 (e-mail: lb247@cornell.edu); Prince: Department of Business Economics and Public Policy, Kelley School of Business, Indiana University, BU460, Bloomington, IN 47405 (e-mail: jeffprin@indiana.edu); Teitelbaum: Georgetown University Law Center, 600 New Jersey Avenue NW, Washington, DC 20001 (e-mail: jct48@law.georgetown.edu). We thank the referees, Dan Benjamin, Larry Blume, Richard Blundell, Keith Chen, Stefano DellaVigna, David Easley, Ani Guerdjikova, Rich Hynes, Howell Jackson, Justin Johnson, Ulrike Malmendier, Matthew Rabin, Julio Rotemberg, Steve Salop, Sharon Tennyson, Tim Vogelsang, Michael Waldman, and especially Francesca Molinari and Ted O'Donoghue, as well as seminar and conference participants at Cornell University, Michigan State University, Queen's University, Williams College, University of Pennsylvania, Georgetown University, the 2007 NBER Summer Institute, the 2008 Econometric Society North American Summer Meeting, the 2008 Conference on Empirical Legal Studies, and the 2009 American Law and Economics Association Annual Meeting. Barseghyan gratefully acknowledges financial support from the Institute for Social Sciences at Cornell University.

¹As used in this paper, the "context" of a decision refers to all aspects of the choice situation other than the economic fundamentals.

household-specific pricing menus of premium-deductible combinations, households' claim histories, and information about the characteristics of each household.

We use the data to test whether a household's deductible choices across coverage types reflect the same degree of risk aversion. Deductible choices are well suited to the task of estimating risk preferences—insurance markets exist primarily because of risk aversion, and the choice of a deductible involves a choice among lotteries over monetary outcomes. Moreover, our dataset is uniquely suited to directly test the assumption that risk preferences are stable across market contexts because it records multiple risky choices by the same households in similar yet distinct market settings. Indeed, it is this key feature of our data that distinguishes us from our predecessors in the literature and allows us to study the issue of context-invariance with market choices.

Our test relies on a structural model of deductible choice that is based on Alma Cohen and Liran Einav (2007). The key assumptions of the model, which we revisit later in the paper, are that households are expected utility maximizers, that claims are generated by a Poisson process at the household-coverage level, and that households know their coverage-specific claim rates and use these claim rates in their expected utility calculus. We impose limited structure on a household's Bernoulli utility function, assuming only that it is smooth, monotone, and state independent and has a negligible third derivative. We also assume that claims always exceed the maximum deductible option, that claims do not entail transaction costs (i.e., costs other than payment of the deductible), and that there is no moral hazard. Under the assumptions of the model, we derive an expression for the degree of absolute risk aversion at which a household is indifferent between two deductible options. This indifference point is a function of the deductible options, the associated premiums, and the household's claim rate. We observe all of the variables necessary to calculate a household's indifference points except for its claim rates. We use data on claim realizations to predict the households' claim rates. In particular, we treat the claim rates as latent random variables and estimate their conditional distributions under the assumption that they depend on observable and unobservable household characteristics.

Our test leverages the idea that each deductible choice by a household implies that its coefficient of absolute risk aversion lies between two indifference points. For example, if the household chooses a deductible of \$500 from a menu of \$100, \$500, and \$1,000, then the indifference point between \$500 and \$1,000 provides a lower bound on its coefficient of absolute risk aversion and the indifference point between \$100 and \$500 provides an upper bound. In this fashion, a household's deductible choices in auto collision, auto comprehensive, and home identify three "test" intervals that must contain its coefficient of absolute risk aversion. If these test intervals fail to intersect, then the household's choices cannot be rationalized by the same coefficient of absolute risk aversion.

It is noteworthy that our test does not rely on statements about the reasonableness of the magnitude of the degree of risk aversion implied by households' choices or on assumptions about the relationship between risk aversion over moderate and large stakes or between risk aversion and wealth. Rather, we test whether a household's market choices over money lotteries of comparable size reflect the same degree of risk aversion measured at a particular wealth level (whatever it may be). Thus, our

study complements prior research on the empirical validity of the standard account of risk preferences (e.g., Paul A. Samuelson 1963; Matthew Rabin 2000; Rabin and Richard H. Thaler 2001).

We find that the hypothesis of stable risk preferences is rejected by the data. Under the benchmark test, which maintains all our modeling assumptions and uses households' predicted claim rates to calculate their test intervals, the fraction of households with threewise success (i.e., for whom the threewise intersection of its test intervals is nonempty) is 23 percent. This is quite low considering that it would be 14 percent if households were randomly assigned their deductible choices. However, one should not expect a success rate of 100 percent, as households' true claim rates are bound to differ from their predicted claim rates due to unobserved heterogeneity. As a relevant point of comparison, we construct the "expected" success rate—i.e., the success rate that we would expect under the null hypothesis of stable risk preferences given our modeling assumptions—and we find that it is 50 percent, more than twice the actual success rate. Furthermore, we show that the residual variance in the claims data cannot likely account for the gap between the predicted claim rates and the claim rates that would be required to rationalize the deductible choices for many households. On the basis of our analysis, we conclude that unobserved heterogeneity is not a plausible explanation for the low rate of success under the benchmark test.

Our analysis of the patterns of failure suggests that households with threewise failure typically exhibit greater risk aversion in their home deductible choices than in their auto deductible choices. Overall, the average failing household would pay more than \$45 to avoid facing a "home" lottery offering an equal chance of winning and losing \$100, but would pay \$30 or less to avoid facing an equivalent "auto" lottery. Our analysis also suggests that, once we control for risk type, none of age, gender, or wealth is predictive of success under the benchmark test.

As a supplement to our primary test, we offer an alternative test that uses the familiar Wald statistic to test whether the empirical distribution of deductible choices in the data is close to the distribution generated by the model under the hypothesis of stable risk preferences. The results of the alternative test reinforce our main findings. In addition, we check the robustness of our benchmark results to a number of alternative modeling assumptions. For instance, we consider the case of constant absolute risk aversion (CARA) utility and we allow for the possibility that households weight their claim rates in the manner suggested by cumulative prospect theory (Amos Tversky and Daniel Kahneman 1992). We also address certain methodological concerns and discuss and defend several other assumptions of the analysis. Finally, we briefly speculate about the potential for three alternative explanations—subjective beliefs, mistakes, and stochastic preferences—to explain our results.

The paper is related to two literatures that cut across economics and psychology. The first estimates risk preferences from observed choices. The majority of the studies in this literature rely on data from surveys (e.g., W. Kip Viscusi and William N. Evans 1990; Evans and Viscusi 1991; Robert J. Barksy et al. 1997; Miles S. Kimball, Claudia R. Sahn, and Matthew D. Shapiro 2008, 2009; Thomas Dohmen et al., forthcoming), laboratory or field experiments (e.g., Steven J. Kachelmeier and Mohamed Shehata 1992; Charles A. Holt and Susan K. Laury 2002; Syngjoo Choi

et al. 2007), or natural experiments in nonmarket settings such as game shows (for reviews, see Pavlo Blavatsky and Ganna Pogrebna 2008; Thierry Post et al. 2008) or horse racing (e.g., Bruno Jullien and Bernard Salanié 2000). Only a handful of papers utilize data on risky choices by market participants. For instance, Atanu Saha (1997) uses data on production decisions to estimate firms' risk preferences and Raj Chetty (2006) estimates risk aversion using data on labor supply decisions. A small group of papers use data on insurance choices. Charles J. Cicchetti and Jeffrey A. Dubin (1994) use the choice of whether to purchase interior telephone wire insurance to estimate a consumer's degree of risk aversion. Cohen and Einav (2007) structurally estimate risk preferences from the deductible choice in auto insurance. Most recently, Justin Sydnor (2010) uses the deductible choice in home insurance to estimate lower and upper bounds for a customer's level of risk aversion. Each of these papers, however, studies unicontext choice and focuses on estimating the magnitude of risk aversion in a single context, whereas we study multicontext choice and focus on the stability of risk preferences across contexts.

The second related literature documents violations of the principle of invariance. According to this principle, preferences over alternatives are invariant to extensionally equivalent reformulations of a decision problem (Tversky and Kahneman 1986). In the case of risky choice, the principle of invariance translates to the consequentialist premise that preferences over risky alternatives are solely a function of the reduced lotteries over outcomes induced by the risky alternatives (Kenneth J. Arrow 1951; Chris Starmer 2000). The studies in the invariance literature generally involve laboratory, field, or natural experiments (for reviews, see Tversky and Thaler 1990; Colin Camerer 1995; Rabin 1998; Starmer 2000; Kahneman 2003a, b).² Notably, several studies find that preferences over money lotteries depend on whether they are framed as a "gamble" or as "insurance" (Paul J. H. Schoemaker and Howard C. Kunreuther 1979; John C. Hershey and Schoemaker 1980; Hershey, Kunreuther, and Schoemaker 1982), and one study finds that preferences over coinsurance depend on whether it is framed as a "deductible" or a "rebate" (Eric J. Johnson et al. 1993). In addition, there are a number of studies that engage in descriptive analysis of certain real world phenomena that appear to contradict invariance (e.g., Thaler 1980; Thomas C. Schelling 1981; Johnson et al. 1993; Edward J. McCaffery 1994; Johnson and Daniel G. Goldstein 2003).

We are aware of only two other studies that present direct evidence on the invariance of risk preferences using data on risky choices by market participants. Charles Wolf and Larry Pohlman (1983) compare the risk preferences of a single dealer in US government securities inferred from his direct assessments of hypothetical gambles and from his bid decisions over a series of Treasury bill auctions. They find that "the dealer was substantially more risk averse in his bid choices than his assessments predicted" and conclude that individuals' "degree of risk aversion may depend on the specific context in which their choices are made" (Wolf and Pohlman 1983, 849). Our paper overcomes two limitations of Wolf and Pohlman (1983). First, we study a large number of households as opposed to a single individual. Second, we study actual choices made in three comparable market contexts,

²Recent examples include Choi et al. (2007) and Lisa R. Anderson and Jennifer M. Mellor (2009).

whereas they compare hypothetical nonmarket choices with actual choices in one market context. Einav et al. (2010) use data on choices by Alcoa employees regarding their 401(k) plans and five employer-provided insurance plans to investigate the stability of risk preferences across contexts. Their results reject the null hypothesis that there is no domain-general component of risk preferences. However, they also find that evidence that risk preferences become more context specific as the choice domains move farther apart. The key difference between Einav et al. (2010) and our paper is that we test a stronger hypothesis. Whereas they test the null hypothesis that risk preferences have no domain-general component, we test the null hypothesis that risk preferences are domain general. We share the view that the two papers (and their results) are highly complementary.

The remainder of the paper proceeds as follows. Section I develops the test of stable risk preferences. It presents the model of deductible choice, explains the construction of the test intervals, and provides the basic intuition for the test. Section II describes the data. Section III delineates the empirical strategy. It discusses the empirical approach to predicting households' claim rates and formulates the benchmark test. Section IV reports the claim rate estimates and the benchmark test results. It also includes our analysis of unobserved heterogeneity and the anatomy of failure. Section V presents the alternative test and a variety of other robustness checks. Concluding remarks appear in Section VI. Supplemental material appears in an Appendix and a Web Appendix.

I. A Test of Stable Risk Preferences

A. Model of Deductible Choice

We model households' deductible choices according to expected utility theory. Let w_i and $u_i(\cdot)$ denote household i 's wealth and Bernoulli utility function, respectively. We assume that $u_i(\cdot)$ is thrice continuously differentiable, monotone, and state independent and has a negligible third derivative. We otherwise impose no structure on the Bernoulli utility function.

Households have access to three types of insurance coverage: auto collision (L), auto comprehensive (M), and home all perils (H).³ Each type of coverage provides full insurance against covered losses in excess of a deductible chosen by the household. The deductible for each coverage is chosen from a household-coverage specific pricing menu that associates a premium with each deductible. For each household i and coverage j , let \mathcal{M}_{ij} denote the pricing menu, d_{ij} denote the deductible, p_{ij} denote the associated premium, and t_{ij} denote the duration of coverage. A policy is a premium-deductible pair, (p_{ij}, d_{ij}) . In the data, auto policies are semiannual contracts and home policies are annual contracts. Accordingly, t_{ij} is the number of semiannual periods of coverage for $j = L, M$ and the number of annual periods of coverage for $j = H$.

We assume that household i 's claims under coverage j are generated by a Poisson process having rate λ_{ij} , where λ_{ij} is a semiannual rate for $j = L, M$ and an annual rate

³ A brief description of each type of coverage appears in the Appendix. For simplicity, we often refer to home all perils simply as home.

for $j = H$. Under this assumption, the household's claim arrivals are independent across coverages (as well as within each coverage) even if its claim rates are correlated. In addition, we assume that: (i) households know their claim rates; (ii) the value of every claim under each coverage exceeds the maximum deductible available for such coverage; (iii) the only cost associated with a claim is payment of the deductible; and (iv) deductible choices do not influence claim rates, i.e., households do not suffer from moral hazard. In Section V, we consider relaxing these and other key assumptions of the model.

Under the terms of the policies in the data, the insured generally can change any deductible or cancel any coverage at any time and pay a linearly weighted or prorated premium. On this basis, we assume that households view their deductible choices as short-term commitments. Accordingly, we model deductible choice as a static decision problem.

Given our assumptions, the expected utility to household i from choosing policies $\{(p_{ij}, d_{ij}): j = L, M, H\}$ from pricing menus $\{\mathcal{M}_{ij}: j = L, M, H\}$ is given by

$$(1) \quad V_i((p_{iL}, d_{iL}), (p_{iM}, d_{iM}), (p_{iH}, d_{iH})) = \\ (1 - \lambda_{iL}t_{iL}) \left\{ \begin{aligned} &(1 - \lambda_{iM}t_{iM})[(1 - \lambda_{iH}t_{iH})u_i(\bar{w}_i) + \lambda_{iH}t_{iH}u_i(\bar{w}_i - d_{iH})] \\ &+ \lambda_{iM}t_{iM}[(1 - \lambda_{iH}t_{iH})u_i(\bar{w}_i - d_{iM}) + \lambda_{iH}t_{iH}u_i(\bar{w}_i - d_{iM} - d_{iH})] \end{aligned} \right\}, \\ + \lambda_{iL}t_{iL} \left\{ \begin{aligned} &(1 - \lambda_{iM}t_{iM})[(1 - \lambda_{iH}t_{iH})u_i(\bar{w}_i - d_{iL}) + \lambda_{iH}t_{iH}u_i(\bar{w}_i - d_{iL} - d_{iH})] \\ &+ \lambda_{iM}t_{iM}[(1 - \lambda_{iH}t_{iH})u_i(\bar{w}_i - d_{iL} - d_{iM}) + \lambda_{iH}t_{iH}u_i(\bar{w}_i - d_{iL} - d_{iM} - d_{iH})] \end{aligned} \right\},$$

where $\bar{w}_i = w_i - p_{iL}t_{iL} - p_{iM}t_{iM} - p_{iH}t_{iH}$ is household i 's net wealth.

B. Construction of Test Intervals

Suppose household i can choose between two auto collision policies (p_{iL}^l, d_{iL}^l) and (p_{iL}^h, d_{iL}^h) , with $d_{iL}^h > d_{iL}^l$ and $p_{iL}^l > p_{iL}^h$.⁴ Given (p_{iM}, d_{iM}) and (p_{iH}, d_{iH}) , household i is indifferent between the two policies when

$$V_i((p_{iL}^l, d_{iL}^l), (p_{iM}, d_{iM}), (p_{iH}, d_{iH})) = V_i((p_{iL}^h, d_{iL}^h), (p_{iM}, d_{iM}), (p_{iH}, d_{iH})).$$

Solving this equation for λ_{iL} and taking limits as $t_{ij} \rightarrow 0$ for all j , we have

$$(2) \quad \lambda_{iL} = \frac{(p_{iL}^l - p_{iL}^h)u_i'(w_i)}{u_i(w_i - d_{iL}^l) - u_i(w_i - d_{iL}^h)}$$

or

⁴To be clear, the superscripts on p_{iL} and d_{iL} correspond to the deductible amount, where l denotes the low deductible and h denotes the high deductible. Accordingly, p_{iL}^l denotes the premium associated with the low deductible, d_{iL}^l , and p_{iL}^h denotes the premium associated with the high deductible, d_{iL}^h .

$$(3) \quad \frac{1}{\lambda_{iL}}(p_{iL}^l - p_{iL}^h)u'_i(w_i) = u_i(w_i - d_{iL}^l) - u_i(w_i - d_{iL}^h).$$

Note that over small time intervals the choice of auto collision policy is not influenced by the choice of auto comprehensive or home policy. This is driven by the fact that claim arrivals are independent and the probability of experiencing two claims under one or more coverages is negligible.

Let $r_{iL}^{l,h}$ denote the coefficient of absolute risk aversion (at wealth level w_i) at which the household is indifferent between the two policies. Taking Taylor expansions of both terms on the right-hand side of equation (3), we obtain

$$r_{iL}^{l,h} \approx \frac{\frac{(p_{iL}^l - p_{iL}^h)}{\lambda_{iL}(d_{iL}^h - d_{iL}^l)} - 1}{\frac{1}{2}(d_{iL}^h + d_{iL}^l)}.$$

Let r_i denote the household’s true coefficient of absolute risk aversion. It follows that the household chooses (p_{iL}^l, d_{iL}^l) if $r_i > r_{iL}^{l,h}$ and (p_{iL}^h, d_{iL}^h) if $r_i < r_{iL}^{l,h}$.

Now suppose household i can choose among K auto collision policies $\{(p_{iL}^k, d_{iL}^k): k = 1, \dots, K\}$, where $d_{iL}^1 < d_{iL}^2 < \dots < d_{iL}^{K-1} < d_{iL}^K$ and $p_{iL}^1 > p_{iL}^2 > \dots > p_{iL}^{K-1} > p_{iL}^K$. When the household chooses policy (p_{iL}^k, d_{iL}^k) , its choice implies

$$r_i \in \mathcal{I}_{iL} = \begin{cases} \left[\max_{h>k} r_{iL}^{1,h}, +\infty \right) & \text{if } k = 1 \\ \left[\max_{h>k} r_{iL}^{k,h}, \min_{l<k} r_{iL}^{l,k} \right] & \text{if } 2 \leq k \leq K - 1 \\ \left(-\infty, \min_{l<K} r_{iL}^{l,K} \right] & \text{if } k = K \end{cases}$$

That is, the household’s choice of auto collision deductible identifies an interval, \mathcal{I}_{iL} , that contains its coefficient of absolute risk aversion. Exactly in the same fashion, the household’s choice of auto comprehensive deductible identifies a second interval, \mathcal{I}_{iM} , and its choice of home deductible identifies a third interval, \mathcal{I}_{iH} . We refer to \mathcal{I}_{iL} , \mathcal{I}_{iM} , and \mathcal{I}_{iH} as household i ’s test intervals.

C. Basic Intuition of the Test

Provided that household i ’s test intervals are nonempty,⁵ we can ask whether they intersect. If the test intervals intersect, then any coefficient of absolute risk aversion contained in the intersection can rationalize the household’s deductible choices. If, however, the test intervals do not intersect, then the household’s deductible choices cannot be rationalized by the same coefficient of absolute risk aversion. We refer to the fraction of households with nonempty intervals/intersections as the success rate. All of the variables necessary to calculate household i ’s test intervals are observable, except for the claim rates, λ_{iL} , λ_{iM} , and λ_{iH} . For this reason, we must use predicted claim rates to calculate households’ test intervals. Because we use predicted claim

⁵For a discussion of empty test intervals, see the Web Appendix.

rates, the expected success rate (i.e., the success rate that we would expect under the null hypothesis of stable risk preferences given our modeling assumptions) is less than 100 percent. Therefore, the conclusion of the test—namely, whether the null hypothesis of stable risk preferences is rejected by the data—turns on comparing the success rate with the expected success rate. We provide a precise formulation of the test, including the procedures for predicting the claim rates and for constructing the expected success rate, in Section III.

II. Data

A. Overview

We acquired the data from an independent insurance agent located in a northeastern state of the United States. The agent offers multiple lines of insurance on behalf of a number of regional and national insurance companies. The data comprise information on households who purchased or renewed auto collision, auto comprehensive, or home coverage through the agent between 1993 and 2007. For each household, the data contain all the information in the agent's records regarding the household's policies, the claims reported thereunder, and the characteristics of the household. The data also include household-coverage specific pricing menus of premium-deductible combinations that we constructed using rating information provided by the agent (see Section IIC).

For purposes of our analysis, we group the data into three nonmutually exclusive samples: Auto, Home, and Test. We use the Auto and Home samples to perform claim rate regressions, the results of which we use to generate predicted claim rates for the Test sample (see Sections IIIA and IVA).⁶ The Auto sample contains 1,689 households who purchased or renewed auto collision or auto comprehensive coverage and for whom we have complete information regarding their claims and the variables that enter as covariates in the claim rate regressions for auto collision or auto comprehensive. Similarly, the Home sample contains 1,298 households who purchased or renewed home coverage and for whom we have complete information regarding their claims and the variables that enter as covariates in the claim rate regression for home. We perform our test of stable risk preferences on the Test sample. The Test sample contains 702 households who purchased or renewed all three types of coverages—auto collision, auto comprehensive, and home—through the agent and for whom we have complete information regarding their deductible choices and the variables necessary to calculate their test intervals. Due to these stringent inclusion requirements, the Test sample is notably smaller than the regression samples. To address concerns about selection issues and raise confidence that the Test sample constitutes a representative sample of the households in the data, we perform informal and formal checks (see Sections IIB, IVA, and VB).

⁶For this reason we occasionally refer to the Auto and Home samples collectively as the regression samples.

TABLE 1—CLAIMS

	Observations	Mean	Standard deviation	Minimum	Maximum
<i>Auto sample</i>					
Collision claim rate (semiannual)	1,689	0.058	0.136	0.000	2.645
Comprehensive claim rate (semiannual)	1,689	0.037	0.092	0.000	1.630
Policy duration (semiannual periods)	1,689	9.79	6.75	0.04	35.46
<i>Home sample</i>					
Claim rate (annual)	1,298	0.052	0.131	0.000	1.848
Policy duration (annual periods)	1,298	5.04	3.29	0.03	16.48

B. Household Characteristics and Claims

Descriptive statistics for the Auto and Home samples, as well as comparable statistics for the Test sample, are set forth in the Web Appendix. An informal comparison of means suggests that moving from the Auto and Home samples to the Test sample does not introduce any worrisome selection bias. This is confirmed when we formally compare the empirical distributions of the predicted claim rates in the Auto and Home samples, on the one hand, and the Test sample, on the other hand (see Section IVA), and when we compare the benchmark test results for the Test sample with test results for the subsample of Test sample households for whom we have complete claims information (see Section VB).

Table 1 summarizes the claim histories of the households in the Auto and Home samples. The mean semiannual claim rates for auto collision and auto comprehensive are 0.058 and 0.037, respectively, and the mean annual claim rate for home is 0.052. We note that our claim rates comport with the national claim rates reported by the Insurance Information Institute (2010a, b).⁷ In addition, we note that: (i) the auto claim rate reported by Cohen and Einav (2007) for a sample of Israeli drivers is greater than the sum of our auto claim rates,⁸ which is consistent with the fact that the road accident and motor vehicle theft rates in Israel are greater than the rates in the United States;⁹ and (ii) the home claim rate reported by Sydnor (2010) for a sample of homeowners in a western state is approximately four-fifths our home claim rate,¹⁰ which seems reasonable considering regional differences in climate and age of housing stock.

⁷The Insurance Information Institute (2010b) reports an average annual home claim rate of 0.052 for 2005–2007. For auto collision and comprehensive, the Insurance Information Institute (2010a) reports average per vehicle annual claim rates of 0.050 and 0.024 for the same period. Although these are lower than our semiannual auto claim rates, we believe three factors account for the discrepancy: (i) the Institute's auto comprehensive claim rate excludes wind and water losses; (ii) our auto claim rates are per policy, 37 percent of which cover at least two vehicles and 11 percent of which cover at least three vehicles; and (iii) the state we study is one of the most expensive states for auto insurance, suggesting that we should expect our claim rates to be above the national average.

⁸Cohen and Einav (2007) report an average annual claim rate of 0.245 for Israeli auto comprehensive coverage, which is an amalgamation of US auto collision and auto comprehensive coverages.

⁹According to the International Traffic Safety Data and Analysis Group (2009), the rate of injury accidents per one million vehicle kilometers in 2008 was 0.35 in Israel and 0.06 in the United States. According to Sergio Herzog (2002), the annual rate of motor vehicle thefts per one thousand registered motor vehicles between 1994 and 1997 ranged from 19.1 to 28.3 in Israel and from 6.4 to 7.6 in the United States.

¹⁰Sydnor (2010) reports an average annual claim rate of 0.042.

C. Pricing Menus, Premiums, and Deductibles

As a group, the households in the data purchased or renewed their auto and home policies from nine insurance companies. Of the nine companies, one offers only auto policies, four offer only home policies, and four offer both auto and home policies. Each company uses the same procedure to price, or rate, its policies. Suppose household i wishes to obtain a quote for coverage j . First, upon observing the household's characteristics, \mathbf{x}_{ij} , the company determines a benchmark premium \bar{p}_{ij} (i.e., the premium associated with a benchmark deductible \bar{d}_j , e.g., $\bar{d}_L = \$200$) according to a proprietary rating function, $\bar{p}_{ij} = f_j(\mathbf{x}_{ij})$. Second, it generates a household-coverage specific pricing menu $\mathcal{M}_{ij} = \{(p_{ij}^k, d_{ij}^k): k = 1, \dots, K\}$ that associates a premium p_{ij}^k with each deductible d_{ij}^k according to a proprietary multiplication rule, $p_{ij}^k = g_j(d_{ij}^k) \bar{p}_{ij}$, with $g_j(\bar{d}_j) = 1$. After the company rates the policy, it quotes the pricing menu to the household. The household then chooses its preferred policy from the pricing menu, $(p_{ij}^*, d_{ij}^*) \in \mathcal{M}_{ij}$.

For each household i in the Auto and Test samples, we observe the household's most recent policy choices $\{(p_{ij}^*, d_{ij}^*): j = L, M, H\}$, but we do not observe the pricing menus $\{\mathcal{M}_{ij}: j = L, M, H\}$ that were quoted to the household. However, we were able to ascertain for all nine companies the multiplication rules $\{g_j: j = L, M, H\}$ that were in effect in 2007. Applying these multiplication rules, we construct pricing menus $\{\hat{\mathcal{M}}_{ij}: j = L, M, H\}$ for each household i in the Auto and Test samples. (We also observe the policy choices of most households in the Home sample. However, unless a household is also in the Auto or Test sample, we do not need to construct pricing menus for the household because we do not use the benchmark premium as a covariate in the home regression (see the Web Appendix)). We use these pricing menus throughout our analysis. In particular, we use them to calculate the benchmark premiums \bar{p}_{iL} and \bar{p}_{iM} , which enter as covariates in the auto regressions, and the test intervals I_{iL} , I_{iM} , and I_{iH} , which enter our test of stable risk preferences.¹¹

We recognize, of course, that $\hat{\mathcal{M}}_{ij} \neq \mathcal{M}_{ij}$ if a different benchmark premium or multiplication rule was in effect at the time household i made its deductible choice for coverage j . However, there are (at least) two reasons to believe that $\hat{\mathcal{M}}_{ij} \cong \mathcal{M}_{ij}$ for nearly all i, j . First, 93 percent of the policies in the Auto sample and 66 percent of the policies in the Test sample were purchased or renewed in 2007. We are confident that the pricing menus for these policies are accurate. Second, each company's rating scheme must be approved by the state insurance department. As a result, the pricing menus are quite stable over time.¹² Nearly 100 percent of the policies in the Auto sample and 89 percent of the policies in the Test sample were purchased or renewed during or after 2004. Therefore, we are confident that the vast majority of the pricing menus are accurate or approximately so.

¹¹Note that $\hat{\mathcal{M}}_{ij}$ includes only the deductible options offered by the company from which household i purchased or renewed its policy for coverage j . This is conservative—including deductible options offered by other companies could only make the household's test intervals narrower.

¹²We were able to confirm for seven companies that their home multiplication rules have not changed since at least 2003. For two of these companies we were able to confirm that their home multiplication rules have not changed since at least 1999. For the eighth home insurance company, we were able to confirm that its multiplication rule has not changed since at least 2006. We also were able to confirm for one company that its auto multiplication rules have not changed since at least 1998. We were not able to trace the evolution of the auto multiplication rules for the other four auto insurance companies.

TABLE 2—PREMIUMS

	Observations	Mean	Standard deviation	Minimum	Maximum
Auto collision premium (dollars)	702	169.56	87.24	27.40	689.00
Auto comprehensive premium (dollars)	702	83.66	51.74	10.90	298.00
Home premium (dollars)	702	580.56	265.27	101.00	2,235.48

TABLE 3—DEDUCTIBLE CHOICES

Deductible	Auto collision		Auto comprehensive		Home	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
\$50	—	—	34	4.84	0	0.00
\$100	9	1.28	28	3.99	16	2.28
\$150	0	0.00	0	0.00	0	0.00
\$200	170	24.22	258	36.75	0	0.00
\$250	127	18.09	100	14.25	487	69.37
\$300	0	0.00	0	0.00	—	—
\$400	0	0.00	0	0.00	—	—
\$500	363	51.71	264	37.61	154	21.94
\$1,000	33	4.70	18	2.56	41	5.84
\$1,500	0	0.00	0	0.00	—	—
\$2,000	0	0.00	0	0.00	1	0.14
\$2,500	0	0.00	0	0.00	3	0.43
\$5,000	0	0.00	0	0.00	0	0.00
\$10,000	0	0.00	0	0.00	0	0.00
Total	702	100.00	702	100.00	702	100.00

Notes: A dash indicates that no company offers the deductible option. Italics indicate that only one company offers the deductible option.

Tables 2 and 3 summarize the policy choices for households in the Test sample. The most popular deductible choices were \$200, \$250, and \$500 for auto collision and auto comprehensive and \$250 and \$500 for home coverage. No household chose an auto deductible in excess of \$1,000, and only four households chose a home deductible in excess \$1,000. In light of this pattern, and taking into consideration the variation in the deductible options across companies, we limit the pricing menus to the following deductible options:¹³

- Auto collision: {\$100, \$200, \$250, \$500, \$1,000, \$2,000, \$2,500};
- Auto comprehensive: {\$50, \$100, \$200, \$250, \$500, \$1,000, \$1,500, \$2,000};
- Home: {\$50, \$100, \$200, \$250, \$500, \$1,000, \$2,000, \$2,500, \$5000}.

Given how we calculate a household's test intervals, limiting the pricing menus to these deductible options is without loss of generality. Indeed, it is conservative because including additional deductible options could only make a household's test intervals narrower.

¹³Of course, each household-coverage specific pricing menu contains only those deductible options that are offered by the company that issued the policy to the household.

III. Empirical Strategy

A. Predicting Claim Rates

All of the variables necessary to calculate household i 's test intervals are observable, except for the claim rates, λ_{iL} , λ_{iM} , and λ_{iH} . We treat the claim rates as latent random variables and assume that for each household i and coverage j ,

$$\ln \lambda_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta}_j + \varepsilon_{ij},$$

where (i) \mathbf{x}_{ij} is a vector of observables and (ii) for each coverage j , the ε_{ij} , which reflect potential unobserved heterogeneity, are independent across households and $\exp(\varepsilon_{ij})$ follows a gamma distribution with unit mean and variance α_j . Under this assumption, the number of claims that household i realizes under coverage j follows a negative binomial distribution with conditional mean $\mu_{ij} = \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) t_{ij}$ and variance $\mu_{ij} + \alpha_j \mu_{ij}^2$ (A. Colin Cameron and Pravin K. Trivedi 1998). Consequently, we can perform standard negative binomial regressions to obtain maximum likelihood estimates of $\boldsymbol{\beta}_j$ and α_j for each coverage j . We then use these estimates to predict the claim rates for each household.

B. Formulation of the Benchmark Test

In the benchmark test, we calculate household i 's test intervals $\{\hat{\mathcal{I}}_{ij}; j = L, M, H\}$ using its predicted claim rates $\{\hat{\lambda}_{ij} = \exp(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}}_j); j = L, M, H\}$. We then verify whether each test interval is nonempty and whether the test intervals' pairwise and threewise intersections are nonempty. Strictly speaking, we verify whether: (i) for each test interval, the upper bound exceeds the lower bound; and (ii) for each pair of test intervals and for the triple $(\hat{\mathcal{I}}_{iL}, \hat{\mathcal{I}}_{iM}, \hat{\mathcal{I}}_{iH})$, the minimum upper bound of the test intervals exceeds the maximum lower bound of the test intervals.

We refer to the fraction of households with nonempty intervals/intersections as the success rate. In general, we do not expect success rates of 100 percent. This is because we calculate the test intervals using households' predicted claim rates as opposed to their true claim rates (which are unobservable). Therefore, we construct expected success rates as the relevant points of comparison. The expected success rate is the success rate that one would expect the sample to achieve in the benchmark test under the null hypothesis that households have stable risk preferences and assuming that our model of deductible choice and our estimates for $\boldsymbol{\beta}_j$ and α_j for each coverage j are correct.¹⁴

We construct the expected success rate as follows. *First*, because we find no unobserved heterogeneity in auto collision (i.e., $\hat{\alpha}_L = 0$; see Section IVA), for each household i we treat its predicted auto collision claim rate, $\hat{\lambda}_{iL}$, as its true auto collision claim rate and we assume that its true coefficient of absolute risk aversion lies in its auto collision interval, $\hat{\mathcal{I}}_{iL}$. *Second*, we take a draw from a uniform distribution

¹⁴For each coverage j , we denote by $\hat{\boldsymbol{\beta}}_j$ and $\hat{\alpha}_j$ our estimates for $\boldsymbol{\beta}_j$ and α_j , respectively.

over $\hat{\mathcal{I}}_{iL}$ and treat it as household i 's true coefficient of absolute risk aversion.¹⁵ *Third*, we draw $\exp(\varepsilon_{iM})$ from $\text{gamma}(1/\hat{\alpha}_M, \hat{\alpha}_M)$ and $\exp(\varepsilon_{iH})$ from $\text{gamma}(1/\hat{\alpha}_H, \hat{\alpha}_H)$ and take $\hat{\lambda}_{iM} \times \exp(\varepsilon_{iM})$ and $\hat{\lambda}_{iH} \times \exp(\varepsilon_{iH})$ as household i 's true auto comprehensive and home claim rates. *Fourth*, we determine each household's optimal deductible choices and calculate the test intervals implied by such choices using its predicted claim rates (just as we do when we compute the benchmark success rate). *Fifth*, we compute the success rate. Repeating steps two through five 1,000 times, we construct the expected success rate (mean) as well as a 95 percent confidence interval.

The conclusion of the test turns on comparing the success rate with the expected success rate. Specifically, we reject the null hypothesis of stable risk preferences if the success rate lies outside the 95 percent confidence interval for the expected success rate.

C. Size and Power of the Test

Because the test is not an exact test in the statistical sense,¹⁶ it is useful to calculate its size and power. Our test rejects the null hypothesis of stable risk preferences if the success rate lies outside the two-sided 95 percent confidence interval for the expected success rate. When the null hypothesis of stable risk preferences is false, the expected success rate can only exceed the success rate. Therefore, under the foregoing rejection rule, the nominal significance level of the test effectively is 2.5 percent. According to our simulations, at the 2.5 percent level the size of the test is 1.7 percent, which suggests that the test is slightly more conservative than the nominal level indicates. This relationship also holds at other common significance levels. The power of the test, which is a function of the magnitude of the deviation of the alternative hypothesis from the null hypothesis of stable risk preferences, has desirable properties as well. At each of the common significance levels, including the 2.5 percent level, we find that the power is low for moderate deviations and then rapidly increases (first at an increasing rate and then at a decreasing rate as it approaches 100 percent) with the magnitude of the deviation. For instance, at the 2.5 percent level, when we allow households' risk aversion coefficients to differ randomly across coverages by no more than 25 percent, our simulations indicate that the power of the test is 2.4 percent; when we allow them to differ by up to 75 percent, the power is 32.9 percent; and when we allow them to differ by as much as 100 percent, the power is 89.5 percent. This implies that relatively small differences in risk preferences across contexts are not likely to trigger rejection of the null hypothesis, suggesting another sense in which the test is conservative. More detailed size and power calculations and a discussion of our procedures appear in the Appendix.

¹⁵ If $\hat{\mathcal{I}}_{iL}$ lacks a finite lower or upper bound (which is the case for 19 households), we close the interval by imposing a lower bound equal to one-half the minimum finite lower bound in the Test sample (if positive) or twice the minimum finite lower bound in the Test sample (if negative) or an upper bound equal to twice the maximum finite upper bound in the Test sample, as the case may be.

¹⁶ In an exact test, the finite sample distribution of the test statistic under the null hypothesis is known.

IV. Results

A. Regressions and Predicted Claim Rates

The results of the claim rate regressions are set forth in the Web Appendix. For auto comprehensive and home, we report the estimates from negative binomial regressions. For auto collision, however, we report the estimates of a Poisson regression, because a likelihood ratio test fails to reject the null hypothesis of equidispersion (i.e., $\alpha_L = 0$).¹⁷ In each regression, the dependent variable is the number of claims, and we control for variation in exposure (i.e., policy duration). The regression results indicate that auto claim rates (collision and comprehensive) are greater for households with a prior accident or a second vehicle. In addition, collision claim rates are greater for households in which the primary driver is married, the primary or secondary driver is female, or any driver is below age 21. Finally, home claim rates are greater for homes that are not owner occupied and increase with the insured value and age of the home.

For purposes of our analysis, we are not interested in the regression results per se. Rather, we use the estimates from the auto and home regressions to generate predicted claim rates— $\{\hat{\lambda}_{ij} = \exp(\mathbf{x}'_{ij}\hat{\beta}_j): j = L, M, H\}$ —for each household i . Table 4 summarizes the predicted claim rates for households in the Test sample and provides comparable statistics for the regression samples. In the Test sample, the mean predicted semiannual claim rates for auto collision and auto comprehensive are 0.055 and 0.037, respectively, and the mean predicted annual claim rate for home is 0.058. These are nearly identical to the corresponding mean predicted claim rates in the regression samples (0.057, 0.037, and 0.059).¹⁸ Indeed, for each coverage j , a Kolmogorov-Smirnov test fails to reject at the 10 percent level the equality of the empirical distribution of $\hat{\lambda}_{ij}$ in the Test sample and in the corresponding regression sample. In terms of risk profile, therefore, we are comfortable that the Test sample constitutes a representative sample of the households in the data.

B. Benchmark Test Results

Table 5 presents the results of the benchmark test. It reports success rates as well as expected success rates. The results clearly favor rejection of the null hypothesis

¹⁷ Our regression model assumes that the number of claims that household i realizes under coverage j follows a negative binomial distribution with conditional mean $\mu_{ij} = \exp(\mathbf{x}'_{ij}\beta_j)t_{ij}$ and variance $\mu_{ij} + \alpha_j\mu_{ij}^2$ (see Section IIIA). The model reduces to a Poisson model if the dispersion parameter, α_j , equals zero (Cameron and Trivedi 1998). If $\alpha_j = 0$, the Poisson quasi-maximum likelihood estimator is asymptotically more efficient (Jeffrey M. Wooldridge 2002). For each coverage j , we estimate both negative binomial and Poisson models and perform a likelihood ratio test of the null hypothesis of equidispersion ($\alpha_j = 0$) against the alternative of overdispersion ($\alpha_j > 0$). As reported in the Web Appendix, we reject the null hypothesis in the case of auto comprehensive and home. The point estimates for α_M and α_H are 0.2171 and 0.5876, respectively. However, we do not reject the null hypothesis in the case of auto collision—the point estimate for α_L is 0.0028 and the p-value of the test statistic is 0.48. Accordingly, we report the negative binomial estimates for auto comprehensive and home and the Poisson estimates for auto collision. (We note that the Poisson and negative binomial estimates for auto collision are virtually identical.) It is noteworthy that Cohen and Einav (2007)—who study Israeli auto comprehensive coverage, which, as noted above, is an amalgamation of US auto collision and auto comprehensive coverages—report a dispersion coefficient of 0.023, which lies between our estimates for α_L and α_M .

¹⁸ They are also in line with the corresponding mean observed claim rates reported in Table 1 (0.058, 0.037, and 0.052).

TABLE 4—PREDICTED CLAIM RATES

	Observations	Mean	Standard deviation	1st percentile	99th percentile
<i>Test sample</i>					
Auto collision	702	0.055	0.031	0.019	0.151
Auto comprehensive	702	0.037	0.018	0.012	0.092
Home	702	0.058	0.017	0.028	0.101
<i>Auto sample</i>					
Auto collision	1,689	0.057	0.030	0.019	0.157
Auto comprehensive	1,689	0.037	0.017	0.013	0.088
<i>Home sample</i>					
Home	1,298	0.059	0.020	0.026	0.111

Notes: Semiannual claim rates for auto collision and auto comprehensive. Annual claim rates for home.

TABLE 5—BENCHMARK TEST RESULTS

Deductible choice(s)	Success rate	Expected success rate	95 percent confidence interval	
Auto collision	0.89	0.89	0.89	0.89
Auto comprehensive	0.89	0.99	0.98	0.99
Home	1.00	0.99	0.98	1.00
Auto collision and auto comprehensive	0.50	0.72	0.70	0.75
Auto collision and home	0.48	0.68	0.65	0.70
Auto comprehensive and home	0.42	0.76	0.73	0.79
Auto collision and auto comprehensive and home	0.23	0.50	0.47	0.54

of stable risk preferences. The success rates for individual deductible choices are respectable—89 percent for auto collision (this is by construction), 89 percent for auto comprehensive, and 100 percent for home—albeit less than expected in the case of auto comprehensive.¹⁹ However, none of the pairwise success rates exceeds 50 percent and each is well outside the 95 percent confidence interval for the corresponding expected success rate. When we consider all three deductible choices, the success rate is only 23 percent. This is quite low considering not only that the expected threewise success rate is 50 percent (with a 95 percent confidence interval of [0.47, 0.54]) but also that the threewise success rate would be 14 percent if households were randomly assigned their deductible choices.²⁰

¹⁹ Similarly, Cohen and Einav (2007) find that about 10 percent of the individuals in their sample choose a high auto deductible when their predicted claim rate implies a choice of low auto deductible for any level of risk aversion.

²⁰ In calculating the pairwise and threewise success rates and expected success rates, we do not distinguish between failures due to empty intervals and failures due to empty intersections. Given $\lambda_{ij} = \lambda_{ij}$, empty intervals are indicative of making dominated choices, whereas empty intersections are indicative of making inconsistent choices. However, if $\lambda_{ij} \neq \lambda_{ij}$, these interpretations are improper. In any case, pooling these two types of failure does not affect the conclusion of the test. In the subsample of 550 households with three nonempty intervals, the pairwise and threewise success rates remain well below the corresponding expected success rates. For instance, the threewise success rate is 0.30 while the threewise expected threewise success rate is 0.57 (with a 95 percent confidence interval of [0.53, 0.61]).

C. Additional Results

Unobserved Heterogeneity.—Among other things, the benchmark test results imply that unobserved heterogeneity cannot fully account for the low rates of success in the Test sample. In this section, we offer a direct assessment of whether unobserved heterogeneity is a plausible explanation for the low benchmark success rates. Our approach is based on the premise that any unobserved household heterogeneity is captured in the model by the error term, ε_{ij} , the exponential of which is assumed to follow a gamma distribution with unit mean and variance α_j .

Intuitively, our strategy is as follows. For each household i in the Test sample, we find the home claim rate $\tilde{\lambda}_{iH}$ nearest to the household's predicted home claim rate $\hat{\lambda}_{iH}$ such that the household's home test interval intersects with both of its auto test intervals (see Figure 1). In other words, $\tilde{\lambda}_{iH}^b$ is the claim rate nearest to $\hat{\lambda}_{iH}^b$ that has the potential to rationalize the household's deductible choices. We then compute the implied error terms, $\tilde{\varepsilon}_{iH}^b = \ln \tilde{\lambda}_{iH}^b - \ln \hat{\lambda}_{iH}^b$, and assess whether their exponentials plausibly could have come from the error distribution we estimate in the Home regression, $\text{gamma}(1/\hat{\alpha}_H, \hat{\alpha}_H)$. By construction, $\tilde{\varepsilon}_{iH}^b$ represents a conservative metric of the minimum deviation from the predicted claim rate $\hat{\lambda}_{iH}^b$ that is necessary to render the household's deductible choices consistent.²¹

We formally proceed in five steps. *First*, for each household i in the Test sample, we generate 10,000 bootstrap test intervals for auto collision and auto comprehensive, $\{(\hat{T}_{iL}^b, \tilde{T}_{iM}^b)\}_{b=1}^{10,000}$, according to the following procedure. In each bootstrap repetition b , we (i) generate bootstrap estimates $\hat{\beta}_L^b$, $\hat{\beta}_M^b$, $\hat{\beta}_H^b$, and $\hat{\alpha}_M^b$ (we do not bootstrap $\hat{\alpha}_L$ because a likelihood ratio test does not reject the null hypothesis that $\alpha_L = 0$ (see Section IVA)); (ii) draw $\exp(\hat{\varepsilon}_{iM}^b)$ from $\text{gamma}(1/\hat{\alpha}_M^b, \hat{\alpha}_M^b)$; and (iii) use $\hat{\lambda}_{iL}^b = \exp(\mathbf{x}'_{ij}\hat{\beta}_L^b)$ and $\tilde{\lambda}_{iM}^b = \exp(\mathbf{x}'_{ij}\hat{\beta}_M^b + \hat{\varepsilon}_{iM}^b)$ to calculate \hat{T}_{iL}^b and \tilde{T}_{iM}^b . *Second*, we construct 10,000 “minimum distance” claim rates for home, $\{\tilde{\lambda}_{iH}^b\}_{b=1}^{10,000}$. The minimum distance claim rate $\tilde{\lambda}_{iH}^b$ solves the program $\min_{\lambda_{iH}^b \in \Lambda_i^b} |\lambda_{iH}^b - \hat{\lambda}_{iH}^b|$, where $\hat{\lambda}_{iH}^b = \exp(\mathbf{x}'_{ij}\hat{\beta}_H^b)$ is the predicted claim rate and Λ_i^b is the set of all claim rates λ_{iH}^b such that the associated home interval \mathcal{I}_{iH}^b (i.e., the interval \mathcal{I}_{iH}^b calculated using λ_{iH}^b) intersects with both of the bootstrap auto intervals \hat{T}_{iL}^b and \tilde{T}_{iM}^b .²² *Third*, we compute “minimum distance” epsilons for home, $\{\tilde{\varepsilon}_{iH}^b\}_{b=1}^{10,000}$, where $\tilde{\varepsilon}_{iH}^b = \ln \tilde{\lambda}_{iH}^b - \ln \hat{\lambda}_{iH}^b$. *Fourth*, we construct the empirical distribution \tilde{F} of $\exp(\tilde{\varepsilon}_{iH}^b)$. For each bootstrap repetition b , we calculate the order statistics $\tilde{\varepsilon}_{(i)H}^b$ of the $\tilde{\varepsilon}_{iH}^b$ and take the average $\tilde{\varepsilon}_{(i)H} = 1/10,000 \sum_{b=1}^{10,000} \tilde{\varepsilon}_{(i)H}^b$ at each position $[i]$. We then define $\tilde{F}(\exp(\varepsilon)) = 1/702 \sum_{i=1}^{702} 1[\exp(\tilde{\varepsilon}_{(i)H}) \leq \exp(\varepsilon)]$, where $1[\cdot]$ is the indicator function. Taking the fifth and ninety-fifth percentiles at each position (i) , we construct a bootstrap 95 percent confidence interval for \tilde{F} . *Fifth*, we compare \tilde{F} to $\hat{F} := \text{gamma}(1/\hat{\alpha}_H, \hat{\alpha}_H)$, the distribution we estimate from the Home sample.

²¹ Alternatively, we could have constructed in the same fashion “minimum distance” epsilons for auto comprehensive, $\tilde{\varepsilon}_{iM}$, and assessed whether the $\exp(\tilde{\varepsilon}_{iM})$ plausibly could have come from $\text{gamma}(1/0.22, 0.22)$. Because, under either approach, we would perform the bootstrap procedure described below, our analysis and conclusions qualitatively would remain unchanged.

²² Strictly speaking, Λ_i^b is the set of all claim rates λ_{iH}^b such that either (i) the lower bound of \mathcal{I}_{iH}^b is less than the minimum upper bound of \hat{T}_{iL}^b and \tilde{T}_{iM}^b or (ii) the upper bound of \mathcal{I}_{iH}^b is greater than the maximum lower bound of \hat{T}_{iL}^b and \tilde{T}_{iM}^b .

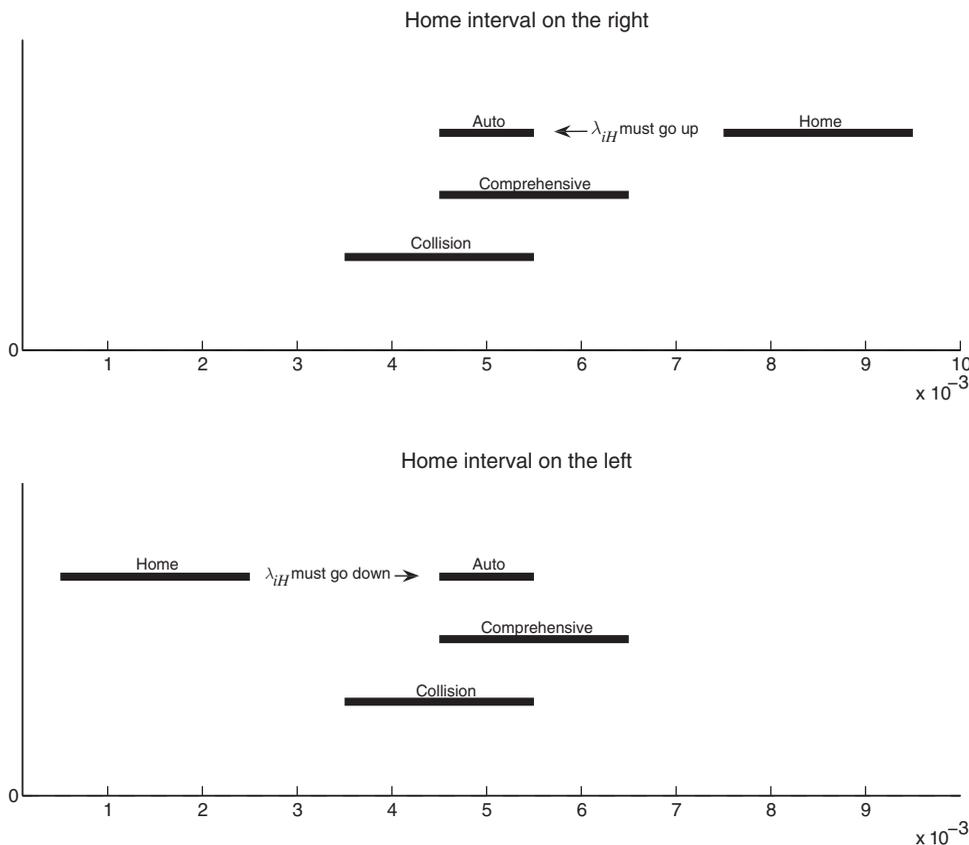
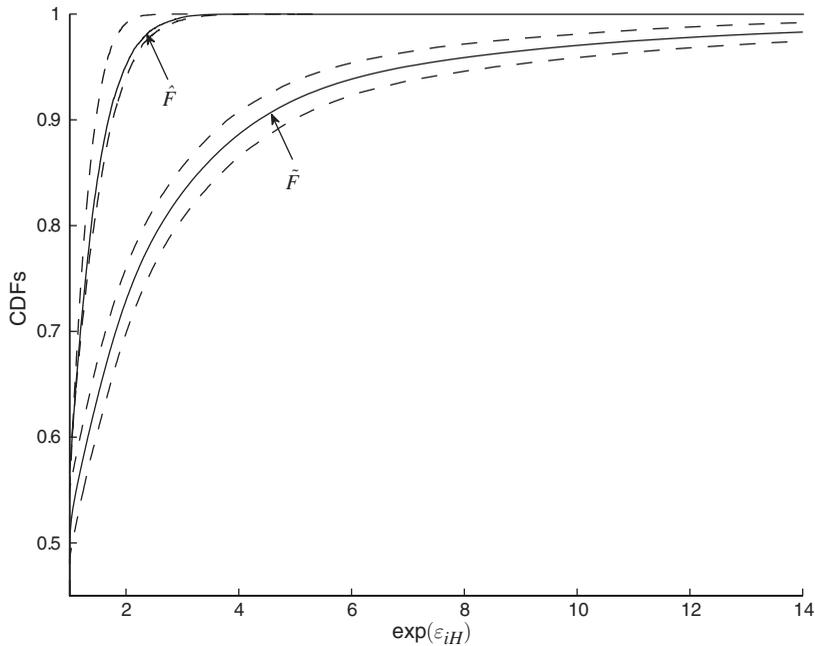


FIGURE 1. CONSTRUCTING “MINIMUM DISTANCE” HOME CLAIM RATES

Figure 2 presents a visual comparison of \tilde{F} and \hat{F} for values of $\exp(\varepsilon_{iH}) \geq 1$.²³ Dashed lines demarcate bootstrap 95 percent confidence intervals.²⁴ It is evident from Figure 2 that the two distributions are not equal. This is formally confirmed by a Kolmogorov-Smirnov test comparing \tilde{F} and the empirical distribution analogously constructed from 10,000 samples of 702 observations drawn from \hat{F} , which rejects at the 0.1 percent level the equality of the distributions for values of $\exp(\varepsilon) \geq 1$. Given the gap between the distributions, it would strain credulity to conclude that unobserved heterogeneity could account for the threewise failure rate under the benchmark test. For instance, according to \tilde{F} , approximately 30 percent of households would require $\exp(\varepsilon_{iH}) \geq 2$ to rationalize their three deductible choices, but according to \hat{F} , the probability of $\exp(\varepsilon_{iH}) \geq 2$ is approximately 5 percent. Moreover, \tilde{F} suggests that at least 10 percent of households would require $\exp(\varepsilon_{iH}) \geq 4$ to rationalize their three deductible choices, but \hat{F} suggest that the probability of $\exp(\varepsilon_{iH}) \geq 4$ essentially is nil. Even if we take a very conservative approach and compare the upper bound of the confidence interval for \tilde{F} with

²³It is not meaningful to compare \tilde{F} and \hat{F} for small values of $\exp(\varepsilon)$ because our procedure for constructing \tilde{F} involves setting $\exp(\varepsilon_{iM}^b) = 0$ if $\hat{\lambda}_{iH}^b \in \Lambda_i^b$, resulting in a point mass at zero.

²⁴The bootstrap 95 percent confidence interval for \tilde{F} reflects the bootstrap 95 percent confidence interval for $\hat{\alpha}_H$.

FIGURE 2. COMPARISON OF \hat{F} AND \tilde{F}

the lower bound of the confidence interval for \hat{F} , the gap appears insurmountable. Accordingly, we conclude that unobserved heterogeneity cannot plausibly explain the low success rates under the benchmark model.

Anatomy of Failure.—For households whose test intervals are empty or fail to intersect, it is natural to ask whether they exhibit any patterns in their characteristics or choices.

There is a clear pattern among the empty test intervals: 70 percent of the empty test intervals in auto collision are attributable to households who purchased or renewed their coverage from Company A, and 68 percent of the empty test intervals in auto comprehensive are attributable to households who purchased their coverage from Company C. What's special about Company A and Company C? In comparison to their peers, Company A has a relatively flat multiplication rule for auto collision, and Company C has a relatively flat multiplication rule for auto comprehensive. Given a deductible choice, the set of claim rates for which the corresponding test interval is nonempty is smaller under a flat multiplication rule than it is under a steep multiplication rule. In the extreme, under a perfectly flat multiplication rule, the set is empty given any “interior” deductible choice. The intuition is straightforward. Under a perfectly flat multiplication rule, it is costless to decrease the deductible and enjoy greater coverage, so whatever its claim rate, a household should choose the minimum deductible option. As the multiplication rule becomes steeper, the costs of decreasing the deductible increase relative to the benefits in terms of greater coverage, so it becomes increasingly worthwhile for households with low claim rates to increase the deductible. Though interesting, the phenomena of empty test intervals

should not be overemphasized. We show below that under CARA utility, the individual success rates increase to nearly 100 percent. We also show that despite such increases in the individual success rates, the pairwise and threewise success rates remain low (see Section VC).

There also are clear patterns in the pairwise position of test intervals for households with three nonempty intervals but one or more empty intersections (384 households). Frequently an auto test interval lies to the left (denoted \triangleleft) of the home test interval: $\hat{\mathcal{I}}_{iL} \triangleleft \hat{\mathcal{I}}_{iH}$ for 54 percent of households and $\hat{\mathcal{I}}_{iM} \triangleleft \hat{\mathcal{I}}_{iH}$ for 63 percent of households. Moreover, when an auto test interval fails to intersect with the home test interval, approximately 85 percent of the time the auto test interval lies to the left of the home test interval. When the auto test intervals fail to intersect (53 percent of households), nearly two-thirds of the time the auto comprehensive test interval lies to the left of the auto collision test interval. For households with three empty intersections (71 households), the modal threewise position is $\hat{\mathcal{I}}_{iM} \triangleleft \hat{\mathcal{I}}_{iL} \triangleleft \hat{\mathcal{I}}_{iH}$ (83 percent of households). These patterns suggest that the typical household with threewise failure exhibits greater risk aversion in its home deductible choice than in its auto deductible choices, and that within auto it exhibits greater risk aversion in its collision deductible choice than in its comprehensive deductible choice.

In an effort to quantify the extent to which households with threewise failure are more risk averse in the home context than in the auto context, we compare for the average household with three nonempty intervals but one or more empty intersections the risk premium of a “home” lottery offering an equal chance of winning and losing \$100 with the risk premium of an equivalent “auto” lottery.²⁵ In our calculations, we assume quadratic utility, under which the risk premium of a lottery p is given by $R(p) = 1/2 \text{Var}(p) r$, where r is the coefficient of absolute risk aversion (Pratt 1964), and we use the average lower bound of the home test intervals and the minimum average upper bound of the auto test intervals as conservative estimates of the average household’s absolute risk aversion coefficients in home and auto, respectively. We find that the average household’s “home” risk premium is \$45 whereas its “auto” risk premium is \$30. This suggests that the typical household with threewise failure is substantially more risk averse in home than it is in auto.

The Web Appendix reports the results of a logit regression in which the dependent variable indicates whether the household’s three test intervals are nonempty and intersect. The results indicate that if we control for risk type, none of age, gender, or wealth is predictive of threewise success under the benchmark test.²⁶

Matching Auto Deductibles.—Nearly 70 percent of the households in the Test sample have matching auto deductibles. One could hypothesize that matching auto deductibles are indicative of a lack of active decision making. An interesting question, therefore, is whether these households have higher or lower success rates than

²⁵ Recall that the risk premium $R(p)$ of a lottery p is the amount such that the household would be indifferent between facing the lottery p and receiving the certainty equivalent $C(p) = E(p) - R(p)$ (John W. Pratt 1964). Stated another way, the risk premium of lottery p is the maximum amount that the household would pay to avoid facing lottery p .

²⁶ We also investigate whether the success rates differ for married and unmarried households. We find that the success rates (as well as the expected success rates) are slightly higher for unmarried households; however, the differences are not material and none of the success rates (or expected success rates) for either subsample materially differs from the success rates (or expected success rates) of the full Test sample.

households without matching auto deductibles. If households with matching auto deductibles have lower success rates, this would be consistent with the hypothesis and might suggest that what we interpret as inconsistent risk preferences is simply inactive decision making. As it turns out, however, the subsample of households with matching auto deductibles (488 households) has a higher pairwise success rate for their auto deductibles (0.52 versus 0.45), as well as a higher threewise success rate (0.25 versus 0.20), than the subsample of households without matching auto deductibles (214 households). This result does not support the hypothesis of inactive decision making.

Implied Estimates of Risk Aversion.—As noted in the introduction, the benchmark test does not rely on statements about the reasonableness of the magnitude of the degree of risk aversion implied by the households' choices. Nevertheless, we can use the households' test intervals—particularly their auto collision test intervals, because there is no unobserved heterogeneity in auto collision (i.e., $\hat{\alpha}_L = 0$)—to estimate the range of the degree of absolute risk aversion of the households in the Test sample. For households with nonempty, bounded auto collision test intervals, the median lower and upper bounds of their auto collision test intervals are 2.0×10^{-3} and 8.2×10^{-3} . For the sake of comparison, we note that Cohen and Einav (2007) report a benchmark estimate of 6.7×10^{-3} for the absolute risk aversion of the mean individual in their data and that Sydnor (2010) reports baseline estimates of 2.0×10^{-3} and 5.3×10^{-3} for the lower and upper bounds on the absolute risk aversion of the median \$500-deductible customer in his data.

V. Discussion and Robustness

A. Alternative Test

The basic premise of the test is that if households have stable risk preferences, then the pairwise and threewise success rates observed in the data should be close to the corresponding success rates generated by the model under the null hypothesis of stable risk preferences. In this section we present an alternative test, the basic premise of which is that if households have stable risk preferences, then, given the households' auto collision deductible choices, the empirical distributions of auto comprehensive and home deductible choices in the data should be close to the corresponding distributions generated by the model under the null hypothesis of stable risk preferences.

For coverage $j = M, H$, let $\phi_j = (\phi_j^1, \dots, \phi_j^K)$ denote the vector of relative frequencies of deductible choices d_j^1, \dots, d_j^K in the data (with N observations) and $\Phi_j = (\Phi_j^1, \dots, \Phi_j^K)$ denote the vector of relative frequencies of deductible choices d_j^1, \dots, d_j^K generated by the model under the null hypothesis of stable risk preferences.²⁷ Although the model-generated distribution, Φ_j , is unique for a given profile of risk preferences $\mathbf{r} = r_1, \dots, r_N$ and claim rates $\boldsymbol{\lambda} = \lambda_{1j}, \dots, \lambda_{Nj}$, we do not directly

²⁷We do not test auto collision because, as indicated above, the households' auto collision deductibles are taken as given. Specifically, as described below, when we generate Φ_j we fix each household's coefficient of absolute risk aversion, r_i , using its auto collision interval.

observe \mathbf{r} and λ . Therefore, we consider 1,000 possible distributions Φ_j , which we construct as follows. For each household i , we set r_i equal to the point in an 11-point uniform grid over its auto collision interval, $\hat{\mathcal{I}}_{iL}$, which maximizes the probability that the household chooses the deductible we observe in the data (auto comprehensive or home, as the case may be).²⁸ We then perform the following routine 1,000 times: (i) for each household i , we draw $\exp(\varepsilon_{ij})$ from $\text{gamma}(1/\hat{\alpha}_j, \hat{\alpha}_j)$ and set $\lambda_{ij} = \hat{\lambda}_{ij} \times \exp(\varepsilon_{ij})$; (ii) we determine each household's optimal deductible choice under the model given r and λ ; and (iii) we compute $\Phi_j^1, \dots, \Phi_j^K$.

The alternative test consists of verifying how often Φ_j belongs to the 99 percent Wald confidence ellipsoid for ϕ_j , which is given by $\mathcal{W}_j = \{\tilde{\Phi}_j: (\tilde{\phi}_j - \tilde{\Phi}_j)' (\widehat{\text{var}}(\tilde{\phi}_j))^{-1}(\tilde{\phi}_j - \tilde{\Phi}_j) \leq c\}$, where $\tilde{\phi}_j = (\phi_j^1, \dots, \phi_j^{K-1})$, $\tilde{\Phi}_j = (\Phi_j^1, \dots, \Phi_j^{K-1})$, $\widehat{\text{var}}(\tilde{\phi}_j)$ denotes the variance-covariance matrix of $\tilde{\phi}_j$, and c is the 99th quantile of the $\chi^2(K-1)$ distribution. If Φ_j lies outside \mathcal{W}_j at least 990 times out of 1,000, then we reject the null hypothesis of stable risk preferences. This rejection rule is very conservative; under it, the size of the test is less than 0.1 percent. Also, the power is well behaved, in the sense that it is low for moderate deviations from the null and high for large deviations (for details, see the Appendix).

The Appendix contains the results of the alternative test for the Test sample (as well as for the various samples and robustness checks considered in Sections VB and VC). When we test the equality of the auto comprehensive distributions, the rejection rate (i.e., the fraction of times that Φ_j lies outside \mathcal{W}_j) is 100 percent for every sample. When we test the equality of the home distributions, the rejection rate is 100 percent for virtually every sample; the only exception is the subsample of Test sample households whose policies were first written or last modified during or after 2006, in which case the rejection rate is 99.5 percent. Thus, the results of the alternative test stoutly reinforce the conclusion of the principal test—the hypothesis of risk preferences is rejected by the data.

The alternative test transparently illustrates the differences between the deductible choices in the data and the model-generated deductible choices. Consider, for example, Figure 3, which compares the joint distribution of auto collision and home deductibles in the Test sample (left panel) with the (average) model-generated joint distribution (right panel).²⁹ Given their auto collision deductibles, the model predicts that the households in the Test sample would choose high home deductibles (\$500 or higher) with greater frequency than we observe in the data. (On average, the model-generated home deductibles differ from the observed home deductibles for at least 263 households, and for at least 183 households the difference equals or exceeds \$750.) This is another way of saying that the households' home deductible choices imply a greater degree of risk aversion than their auto collision deductible choices (which accords with the patterns of failure under the benchmark test (see Section IVC)). The reverse image emerges when we compare the empirical and model-generated joint distributions of auto collision and auto comprehensive deductibles: the model predicts a higher frequency of low auto comprehensive

²⁸Note that using a 5-point grid or a 21-point grid would not change the results.

²⁹To avoid "small" cells problems, we group the auto collision deductibles into four bins (\$100 or lower, \$200 or \$250, \$500, and \$1,000 or higher) and the home deductibles into three bins (\$250 or lower, \$500, and \$1,000 or higher).

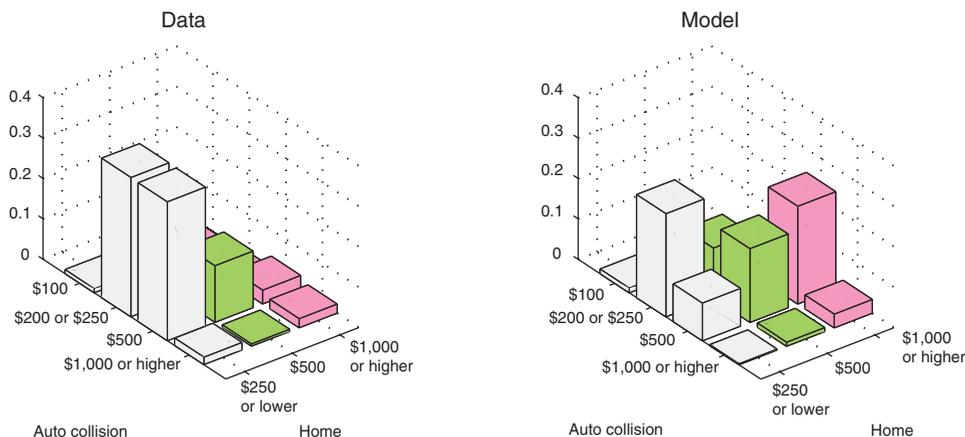


FIGURE 3. COMPARISON OF EMPIRICAL AND MODEL-GENERATED JOINT DISTRIBUTIONS OF AUTO COLLISION AND HOME DEDUCTIBLES

deductibles (\$250 or lower) than we observe in the data, suggesting that the households' auto comprehensive deductible choices imply a lesser degree of risk aversion than their auto collision deductible choices (which also accords with the patterns of failure under the benchmark test). (On average, the model-generated auto comprehensive deductibles differ from the observed auto comprehensive deductibles for at least 309 households.)

B. Methodological Concerns

In this section, we address certain concerns regarding our empirical methodology.

Timing of Purchases.—For each household in the Test sample, we observe its most recent policy for each coverage and we construct its coverage-specific pricing menus by applying the company-specific multiplication rules in effect in 2007 (see Section IIC). Because the majority of the policies in the Test sample are renewal policies, there may be a concern that, insofar as households' deductible choices are “sticky” (i.e., insensitive to changes in premiums), certain of the policy choices we observe are not a function of the pricing menus that we construct, but rather a function of the pricing menus in effect on the date the policy was first written or last actively modified. To address this concern, we run the benchmark test on the subsample of Test sample households whose policies were first written or last modified during or after 2006.³⁰ Table 6 presents the results.³¹ As compared to the benchmark results, both the success rates and the expected success rates increase somewhat,

³⁰The agent's database shows the status of a policy as either new, renewal, changed, or canceled. A changed policy is a policy with one or more changed terms. We consider the effective date of a changed policy as the date on which the policy was last actively modified. Although we cannot observe whether the changes involve the deductible, we consider any change as indicating (or at the very least prompting) active reconsideration of the policy, including the choice of deductible.

³¹It is worth noting that the results barely change if we expand the time frame to include policies first written or last modified during or after 2005, 2004, or 2003.

TABLE 6—SUCCESS RATES: POLICIES FIRST WRITTEN OR LAST MODIFIED DURING OR AFTER 2006

Deductible choice(s)	Observations	Success rate	Expected success rate	95 percent confidence interval	
Auto collision	187	0.93	0.93	0.93	0.93
Auto comprehensive	187	0.89	0.99	0.98	1.00
Home	138	1.00	0.99	0.98	1.00
Auto collision and auto comprehensive	187	0.56	0.77	0.72	0.81
Auto collision and home	72	0.56	0.73	0.65	0.81
Auto comprehensive and home	72	0.56	0.77	0.67	0.86
Auto collision and auto comprehensive and home	72	0.29	0.54	0.43	0.65

with the former increasing somewhat more than the latter. All in all, however, the results do not differ materially from the benchmark results; the pairwise and three-wise success rates remain well below the corresponding expected success rates.³²

Two related concerns are that households first purchased their auto and home policies at different times and that different household members made the auto and home policy choices.³³ To address these concerns, we run the benchmark test on two subsamples: the subsample of 179 Test sample households who first purchased or last modified their auto and home policies within 6 months of each other and the subsample of 201 Test sample households with one member.³⁴ For each subsample, the pairwise and threewise success rates are well below (and outside the confidence interval for) the corresponding expected success rates. The threewise success rates for the subsamples are 26 percent and 28 percent, respectively, whereas the corresponding expected success rates are 50 percent (with a 95 percent confidence interval of [0.43, 0.57]) and 54 percent (with a 95 percent confidence interval of [0.46, 0.62]), respectively.

Construction of Predicted Claim Rates.—We use the Auto and Home samples to perform the regressions described in Section IIIA and then use the results of these regressions to construct the predicted claims rates for the households in the Test sample. The Test sample, however, neither comprises nor is a strict subset of the intersection of the Auto and Home samples.³⁵ Therefore, even though we cannot

³²We also run the benchmark test on the subsample of Auto sample households whose policies were first written or last modified during or after 2006. The Web Appendix contains the results, as well as the benchmark results, for the entire Auto sample for purposes of comparison. The success rates in the Auto sample are substantially the same as the success rates in the Test sample. Moreover, restricting the Auto sample to households with policies first written or last modified during or after 2006 does not materially alter the results—the success rates remain well below the corresponding expected success rates.

³³Two recent papers that examine the stability of risk preferences over time are Steffen Andersen et al. (2008) and Manel Baucells and Antonio Villasís (2010). One recent paper that studies covariation of risk preferences among household members is Kimball, Sahn, and Shapiro (2009).

³⁴We consider a household to have one member if its auto policy lists only one vehicle and a sole, unmarried driver.

³⁵Of the 702 households in the Test sample, 265 are in the Auto and Home samples, 60 are in the Auto sample but not the Home sample, 201 are in the Home sample but not the Auto sample, and 176 are in neither regression sample. The principal reason why a household in the Test sample is not in a regression sample is because the household has an older policy for which we cannot pinpoint the exact date it was first written. Although we often have complete claims information for the household, because we cannot precisely compute the duration of its policy, we conservatively exclude the household from the regression sample.

reject the equality of the empirical distributions of the predicted claim rates in the Auto and Home samples, on the one hand, and the Test sample, on the other hand (see Section IVA), it nevertheless is possible that the relationship between observable household characteristics and household risk type (claim rate) is different in the Test sample than it is in the regression samples. Although in our opinion there is no reason to believe that there is a correlation between risk type in auto or home and the criteria for inclusion in the Auto, Home, or Test sample, we perform two checks to address this concern. Both involve testing the “intersection subsample”—the subsample of 265 Test sample households that also appear in both regression samples. In the first check, we construct the test intervals using the benchmark claim rates, i.e., the claim rates predicted using the regression results reported in Section IVA. In the second check, we construct the test intervals using “subsample claim rates,” i.e., claim rates predicted from regressions in which we restrict each regression sample to the intersection subsample. The results are set forth in Table 7. Comparing the results to the benchmark test results set forth in Table 5, we find no material differences; on the contrary, they are substantially similar.³⁶

Claims Information.—In the benchmark test, we assume that household i 's claim rate for coverage j is a (stochastic) function of observable household characteristics only. Specifically, we assume $\lambda_{ij} = \exp(\mathbf{x}'_{ij}\hat{\beta}_j) \cdot \exp(\varepsilon_{ij})$, where \mathbf{x}_{ij} is a vector of observable household characteristics and $\exp(\varepsilon_{ij}) \sim \text{gamma}(1/\hat{\alpha}_j, \hat{\alpha}_j)$, which implies $E(\exp(\varepsilon_{ij})) = 1$. We then use $\hat{\lambda}_{ij} = E(\lambda_{ij} | \mathbf{x}_{ij}) = \exp(\mathbf{x}'_{ij}\hat{\beta}_j)$ as household i 's predicted claim rate for coverage $j = L, M, H$ for purposes of computing the test intervals/success rates, and $\hat{\lambda}_{ij} \times \exp(\varepsilon_{ij})$ as its true claim rate for coverage $j = M, H$ for purposes of computing the expected success rates (see Section IIIB). For households for whom we have claims information, however, we could utilize this information to further refine our inference about their claim rates. If z_{ij} denotes the number of claims that household i with characteristics \mathbf{x}_{ij} realizes under coverage j with duration t_{ij} , then

$$\exp(\varepsilon_{ij}) | z_{ij} \sim \text{gamma}\left(\frac{1}{\hat{\alpha}_j} + z_{ij}, \hat{\alpha}_j / (1 + \exp(\mathbf{x}'_{ij}\hat{\beta}_j)t_{ij}\hat{\alpha}_j)\right),$$

which implies

$$E(\exp(\varepsilon_{ij}) | z_{ij}) = \frac{1 + z_{ij}\hat{\alpha}_j}{1 + \exp(\mathbf{x}'_{ij}\hat{\beta}_j)t_{ij}\hat{\alpha}_j}.$$

Were we to modify the test to incorporate households' claims information, therefore, we would assign higher claim rates to households who realize a greater number of claims, all else equal. In particular, we would use

$$\hat{\lambda}_{ij} = E(\lambda_{ij} | \mathbf{x}_{ij}, z_{ij}, t_{ij}) = \exp(\mathbf{x}'_{ij}\hat{\beta}_j) \times \frac{1 + z_{ij}\hat{\alpha}_j}{1 + \exp(\mathbf{x}'_{ij}\hat{\beta}_j)t_{ij}\hat{\alpha}_j}$$

³⁶One noteworthy difference when we use the subsample claim rates is that the expected success rates increase. This is because our estimates for α_j (unobserved heterogeneity) decrease when we restrict each regression to the subsample.

TABLE 7—SUCCESS RATES: INTERSECTION SUBSAMPLE

Deductible choice(s)	Test sample				Intersection subsample			
	Success rate	Expected success rate	95 percent confidence interval		Success rate	Expected success rate	95 percent confidence interval	
Auto collision	0.88	0.88	0.88	0.88	0.91	0.91	0.91	0.91
Auto comprehensive	0.88	0.99	0.98	1.00	0.86	1.00	1.00	1.00
Home	1.00	0.99	0.98	1.00	1.00	1.00	0.99	1.00
Auto coll. and auto comp.	0.53	0.71	0.67	0.75	0.44	0.91	0.91	0.91
Auto coll. and home	0.49	0.68	0.64	0.72	0.54	0.78	0.74	0.81
Auto comp. and home	0.39	0.76	0.71	0.80	0.34	0.91	0.88	0.94
Auto coll. and auto comp. and home	0.23	0.50	0.45	0.55	0.20	0.74	0.70	0.78

TABLE 8—SUCCESS RATES: CLAIMS INFORMATION

Deductible choice(s)	Test sample				Intersection subsample			
	Success rate	Expected success rate	95 percent confidence interval		Success rate	Expected success rate	95 percent confidence interval	
Auto collision	0.89	0.89	0.89	0.89	0.88	0.88	0.88	0.88
Auto comprehensive	0.89	0.99	0.98	0.99	0.88	0.99	0.98	1.00
Home	1.00	0.99	0.99	1.00	1.00	0.99	0.98	1.00
Auto coll. and auto comp.	0.48	0.73	0.71	0.75	0.53	0.72	0.68	0.76
Auto coll. and home	0.48	0.70	0.67	0.72	0.49	0.71	0.67	0.75
Auto comp. and home	0.41	0.77	0.74	0.80	0.36	0.78	0.73	0.83
Auto coll. and auto comp. and home	0.23	0.53	0.49	0.56	0.21	0.53	0.47	0.58

as household i 's predicted claim rate for coverage $j = L, M, H$ for purposes of computing the test intervals/success rates, and $\exp(\mathbf{x}'_{ij}\hat{\beta}_j) \times \exp(\varepsilon_{ij}) | z_{ij}$ as its true claim rate for coverage $j = M, H$ for purposes of computing the expected success rates. As it turns out, however, modifying the test to incorporate claims information does not meaningfully affect the benchmark results. Whether we look at the full Test sample (where we have auto claims information for 452 households and home claims information for 466 households) or the intersection subsample (where we have auto and home claims information for all households), the results barely change when we incorporate claims information (see Table 8).³⁷ Likewise, incorporating claims information does not meaningfully affect any of the other results and robustness checks presented in Sections VB and VC.

C. Model Assumptions

In this section, we explore the robustness of our benchmark test results to alternative modeling assumptions.

³⁷The benchmark results for the intersection subsample are reported in Table 7.

CARA Utility.—The benchmark model assumes that the Bernoulli utility function has a negligible third derivative. Under this assumption, a household's coefficient of absolute risk aversion is a "sufficient statistic" for its risk preferences. However, one might question this assumption, particularly in light of the importance of the third derivative in other literatures in economics and insurance, including the precautionary saving literature (e.g., Kimball 1990) and the life insurance literature (e.g., Nicholas Economides 1982).

Following Cohen and Einav (2007), we check the robustness of the benchmark test results to this assumption by considering the case of CARA utility.³⁸ Under CARA utility, where $u(w) = -\exp(-rw)$, equation (2) becomes

$$\lambda_{ij} = \frac{r_{ij}^{l,h}(p_{ij}^l - p_{ij}^h)}{\exp(r_{ij}^{l,h}d_{ij}^h) - \exp(r_{ij}^{l,h}d_{ij}^l)}.$$

Although this equation does not admit a closed-form solution for $r_{ij}^{l,h}$, a straightforward application of the intermediate value theorem establishes that a unique positive solution exists. Hence, as before, we can construct household i 's test intervals from its deductible choices.

Table 9 reports the success rates and expected success rates under CARA utility.³⁹ The success rates under CARA utility are higher than the benchmark success rates. Indeed, the individual success rates are nearly 100 percent for auto collision and auto comprehensive and reach 100 percent for home. However, the pairwise and threewise success rates remain well below the corresponding expected success rates, which also are higher under CARA utility.

Probability Weighting.—The benchmark model assumes not only that each household knows its objective claim rates but that the decision weights that enter its expected utility calculus correspond to its objective claim rates. We probe the sensitivity of the benchmark test results to this assumption by allowing for the possibility that households distort their claim rates in the manner suggested by cumulative prospect theory (Tversky and Kahneman 1992). Specifically, we assume that households weight their claim rates according to the inverse-S-shaped probability weighting function $\pi_i(\lambda_{ij}) = (\lambda_{ij})^\gamma / [(\lambda_{ij})^\gamma + (1 - \lambda_{ij})^\gamma]^{1/\gamma}$, where $\gamma \leq 1$, and (to preserve monotonicity) that they evaluate their choice of deductible according to the rank dependent value function $V_i(p_{ij}, d_{ij}) = (1 - \pi_i(\lambda_{ij})t_{ij}) \times u(w_i - p_{ij}t_{ij}) + \pi_i(\lambda_{ij})t_{ij} \times u(w_i - p_{ij}t_{ij} - d_{ij})$.⁴⁰

Table 10 reports the success rates and expected success rates for selected values of the parameter γ , including the estimate reported by Tversky and Kahneman

³⁸ As noted by Cohen and Einav (2007), CARA utility allows for a positive third derivative without expanding the dimensionality of preferences (i.e., household i 's preferences still may be summarized by a single parameter, r_i , the household's coefficient of absolute risk aversion) or requiring us to make additional assumptions about households' wealth levels.

³⁹ As in the benchmark test, we calculate household i 's test intervals $\{\hat{x}_{ij}; j = L, M, H\}$ using its predicted claim rates $\{\hat{\lambda}_{ij} = \exp(\mathbf{x}_{ij}\hat{\beta}_j); j = L, M, H\}$.

⁴⁰ For purposes of our analysis in this section, we normalize p_{ij} and λ_{ij} to reflect semiannual premiums and claim rates for each coverage j .

TABLE 9—SUCCESS RATES: CARA UTILITY

Deductible choice(s)	Success rate	Expected success rate	95 percent confidence interval	
Auto collision	0.99	0.99	0.99	0.99
Auto comprehensive	0.98	1.00	1.00	1.00
Home	1.00	1.00	1.00	1.00
Auto collision and auto comprehensive	0.63	0.83	0.80	0.85
Auto collision and home	0.54	0.84	0.81	0.86
Auto comprehensive and home	0.48	0.82	0.79	0.84
Auto collision and auto comprehensive and home	0.30	0.66	0.63	0.69

TABLE 10—SUCCESS RATES: PROBABILITY WEIGHTING

Deductible choice(s)	$\gamma = 0.16$		$\gamma = 0.40$		$\gamma = 0.69$	
	Success rate	Expected success rate	Success rate	Expected success rate	Success rate	Expected success rate
Auto collision	0.97	0.97	0.55	0.55	0.76	0.76
Auto comprehensive	0.89	1.00	0.34	0.98	0.60	0.96
Home	1.00	1.00	0.79	0.98	0.97	0.97
Auto coll. and auto comp.	0.46	0.96	0.12	0.52	0.29	0.62
Auto coll. and home	0.64	0.95	0.33	0.51	0.45	0.63
Auto comp. and home	0.46	0.99	0.12	0.94	0.27	0.79
Auto coll. and auto comp. and home	0.25	0.93	0.05	0.48	0.15	0.49

(1992) ($\gamma = 0.69$).⁴¹ For low values of γ , the threewise success rates with probability weighting are higher than the benchmark threewise success rates. This is because for low values of γ the average household overweights its home claim rate more than it overweights its auto claim rates, which partially accounts for the typical pattern of failure (i.e., greater risk aversion in home than in auto).⁴² Nevertheless, the pairwise and threewise success rates remain well below the corresponding expected success rates, which also increase from their benchmark levels. For moderate and high values of γ , the average household overweights its auto claim rates more than it overweights its home claim rate; as a result, the success rates with probability weighting are lower than the benchmark success rates and remain well below the corresponding expected success rates.

Prospect Theory.—We assume that households make their deductible choices according to expected utility theory. A recent study by Sydnor (2006), however, finds that a variant of cumulative prospect theory (Tversky and Kahneman 1992)—the

⁴¹ In addition to the values of γ reported in Table 10, we consider 0.05, 0.10, 0.20, 0.25, 0.55, and 0.85: the threewise success rates (expected success rates) are 0.30 (0.92), 0.30 (0.93), 0.18 (0.85), 0.09 (0.73), 0.08 (0.46), and 0.21 (0.50), respectively. All together, the results suggest a U-shaped relationship between γ and the success rate.

⁴² Indeed, for $\gamma = 0.16$, the average household overweights its home claim rate and underweights its auto claim rates. For even lower values of γ , the average household underweights all its claim rates, with home underweighted less than auto. The threewise success rate, however, appears to plateau around 30 percent.

so-called “no loss aversion in buying” specification (Nathan Novemsky and Kahneman 2005), in which payments made for purchase (including insurance premiums) are treated as negative gains instead of losses (David H. Krantz and Kunreuther 2007)—does a better job than expected utility theory of explaining observed home deductible choices in data similar to ours. It behooves us, therefore, to verify the robustness of the benchmark success rates to Sydnor’s model. Under his model, household i chooses policy (p_{ij}, d_{ij}) to maximize $V_i(p_{ij}, d_{ij}) = v(-p_{ij}) + \pi(\lambda_{ij}t_{ij}) \times \xi_{ij} \times v(-d_{ij})$, where ξ_{ij} is the coefficient of loss aversion. The functional forms of the value function and the probability weighting function are given by $v(-x) = -(x)^\rho$ and $\pi(x) = (x)^\gamma / [(x)^\gamma + (1-x)^\gamma]^{1/\gamma}$, and the shape parameters are $\rho = 0.88$ and $\gamma = 0.69$. Under this model, the household is indifferent between two policies (p_{ij}^l, d_{ij}^l) and (p_{ij}^h, d_{ij}^h) when

$$\xi_{ij}^{l,h} = \frac{(p_{ij}^l)^\rho - (p_{ij}^h)^\rho}{\pi(\lambda_{ij}t_{ij})[(d_{ij}^h)^\rho - (d_{ij}^l)^\rho]}.$$

Let ξ_i denote the household’s true coefficient of loss aversion. By the same logic set forth in Section IB, when the household chooses policy (p_{ij}^k, d_{ij}^k) from pricing menu $\mathcal{M}_{ij} = \{(p_{ij}^k, d_{ij}^k): k = 1, \dots, K\}$, its choice identifies an interval, \mathcal{I}_{ij} , that contains its coefficient of loss aversion, and we can check whether its test intervals $\{\mathcal{I}_{ij}: j = L, M, H\}$ are nonempty and intersect. The Web Appendix reports the success rates and expected success rates under Sydnor’s prospect theory model. The pairwise and threewise success rates are uniformly lower under Sydnor’s model and they remain far below the corresponding expected success rates.

Gamma Errors.—In the model, we assume that $\ln \lambda_{ij} = \mathbf{x}'_{ij}\beta_j + \varepsilon_{ij}$, where $\exp(\varepsilon_{ij})$ follows a gamma distribution with unit mean and variance α_j . Under this assumption, the number of claims that household i realizes under coverage j follows a negative binomial distribution and, therefore, we can estimate β_j and α_j using a negative binomial regression model. In addition to being mathematically convenient, the assumption that $\exp(\varepsilon_{ij})$ follows a gamma distribution is motivated by the fact that the negative binomial distribution historically has provided a good fit for accident data (Major Greenwood and G. Udny Yule 1920; Georges Dionne and Charles Vanasse 1992). There are many examples of negative binomial regression models in the economics and insurance literatures (e.g., Jerry Hausman, Bronwyn H. Hall, and Zvi Griliches 1984; Cameron and Trivedi 1986; Marcel Boyer, Dionne, and Vanasse 1992; Dionne and Vanasse 1992). Even if $\exp(\varepsilon_{ij})$ does not follow a gamma distribution, the negative binomial quasi-maximum likelihood estimator for β_j is consistent under the conditional mean assumption $E(\text{claims}_{ij} | \mathbf{x}_{ij}) = \exp(\mathbf{x}'_{ij}\beta_j)t_{ij}$ only; however, the estimator for α_j generally will be inconsistent if there is any distributional misspecification (Cameron and Trivedi 1998). Therefore, it is important to check the sensitivity of our error variance estimates to this distributional assumption. Under the conditional variance assumption $\text{var}(\text{claims}_{ij} | \mathbf{x}_i) = \mu_{ij} + \alpha_j\mu_{ij}^2$, it is possible to obtain a distribution-free consistent estimator for the error variance given an initial estimator for β_j (Wooldridge 2002). We use the following strongly

TABLE 11—ERROR VARIANCE ESTIMATES

	$\hat{\alpha}_j$	$\hat{\eta}_j$	Bootstrap 95 percent confidence interval	
Auto collision ($j = L$)	0.00	0.07	-0.13	0.19
Auto comprehensive ($j = M$)	0.22	0.20	-0.12	0.40
Home ($j = H$)	0.59	0.51	-0.01	1.57

consistent estimator suggested by Christian Gouriéroux, Alain Monfort, and Alain Trognon (1984):⁴³

$$\hat{\eta}_j = \frac{\sum_{i=1}^N [(claims_{ij} - \exp(\mathbf{x}'_{ij}\hat{\beta}_j))^2 - \exp(\mathbf{x}'_{ij}\hat{\beta}_j)] \exp(2\mathbf{x}'_{ij}\hat{\beta}_j)}{\sum_{i=1}^N \exp(4\mathbf{x}'_{ij}\hat{\beta}_j)}$$

Table 11 compares the point estimates that result from our benchmark negative binomial estimator $\hat{\alpha}_j$ with the point estimates that result from the strongly consistent estimator $\hat{\eta}_j$. The benchmark estimates are slightly higher than the strongly consistent estimates in the case of auto comprehensive and home and slightly lower in the case of auto collision. In each case, however, the magnitude of the difference is small and the benchmark estimate lies within the 95 percent confidence interval for the strongly consistent estimate.

Notwithstanding the foregoing, one still might question whether our analysis and conclusions in Section IVC regarding unobserved heterogeneity are robust to alternative distributional assumptions. Of course, any sensible alternative error distribution must have unit mean and support on the nonnegative real numbers. In addition, per the above, the variance of any alternative error distribution should be approximately equal to the relevant benchmark estimate in order to be faithful to the data. Lastly, Figure 2 suggests that any viable alternative error distribution (i.e., any distribution that could rationalize the data) would have to have an extremely heavy tail. There is reason to believe, however, that no sensible alternative error distribution could have a sufficiently heavy tail to be viable. This is because Figure 2 suggests that the tail of any viable alternative error distribution would have to be so heavy that it could not have unit mean. According to \tilde{F} (the empirical distribution of $\exp(\tilde{\varepsilon}_{iH})$, the exponential of the “minimum distance” epsilons for home), $E(\exp(\tilde{\varepsilon}_{iH}) | \exp(\tilde{\varepsilon}_{iH}) > 1) = 6.83$ and $\Pr(\exp(\tilde{\varepsilon}_{iH}) > 1) = 0.56$, which implies $E(\exp(\tilde{\varepsilon}_{iH})) > 1$. Accordingly, we believe that our conclusion regarding the inability of unobserved heterogeneity to explain the low benchmark success rates does not depend on our assumption of gamma distributed errors.

Claim Size.—The model assumes that every claim exceeds the maximum deductible option for the relevant coverage. Unfortunately, we cannot confirm this assumption for our data because we do not observe claim amounts. However, reports from

⁴³Note that $\hat{\beta}_j$ denotes our benchmark negative binomial estimator for β_j .

studies using similar data support this assumption.⁴⁴ Furthermore, as long as every claim exceeds \$500, our benchmark test intervals are conservative for the overwhelming majority of households in the Test sample. To see this, consider the following example. Suppose that there are two deductible options, $d^h > d^l$, and that all claims exceed d^l , but not d^h . As before, we can derive an expression for the coefficient of absolute risk aversion at which the household is indifferent between the two deductible choices. Let $\tilde{r}^{l,h}$ denote this coefficient. Note that $\tilde{r}^{l,h} > r^{l,h}$ (the coefficient of absolute risk aversion at which the household is indifferent between the two policies under the benchmark assumptions). The intuition is as follows. The high deductible, d^h , is more attractive to the household in the case in which some claim amounts are less than d^h than it is in the benchmark case in which all claims amounts exceed d^h . This is because in the benchmark case all claims result in a loss equal to d^h , whereas in the former case some claims result in a loss less than d^h . Consequently, the degree of risk aversion required for the household to choose the low deductible, d^l , is greater in the benchmark case, which implies $\tilde{r}^{l,h} > r^{l,h}$. In general, this example shows that if our claim size assumption were incorrect, then the lower and upper bounds of the household's test interval would be too small. However, if all claims exceed \$500, this means that our benchmark test intervals are conservative for the overwhelming majority of the households in our sample: (i) for households with a deductible less than \$500, the benchmark test intervals are correct; (ii) for households with a \$500 deductible, the lower bound of the benchmark test interval is too low but the upper bound is correct, which implies that the benchmark test interval covers the true test interval; and (iii) for households with a deductible in excess of \$500, although the benchmark interval does not fully cover the true test interval, the fraction of such households in our sample is small (the fraction of households with a deductible in excess of \$500 is less than 5 percent in auto collision, less than 3 percent in auto comprehensive, and approximately 6 percent in home). In fact, the three-wise success rate and expected success rate for the subsample of households with deductibles all less than \$1,000 (23.9 percent and 50.4 percent) are nearly identical to the success rate and expected success rates for the full Test sample (23.4 percent and 49.6 percent).

Transaction Costs.—The model assumes that the only cost associated with a claim is payment of the deductible. In reality, however, there may be other transaction costs associated with filing and pursuing a claim. Abstracting from such transaction costs, particularly to the extent that they are greater for home claims than they are for auto claims, could bias the results of the benchmark test. As a check of the robustness of the benchmark test results to this assumption, we allow for the possibility that home claims involve positive transaction costs, ranging from \$100 to \$500 in \$100 increments. We find that inclusion of such transaction costs does not alter the results in any material respect; the relevant pairwise success rates increase

⁴⁴ Using auto data similar to ours with two deductible options, Cohen and Einav (2007) report that 99 percent of the claims filed by those insured with the low deductible were for amounts greater than the high deductible option. Using home data similar to ours with four deductible options, Sydnor (2010) reports that the average claim at each deductible level was greater than the highest deductible option. In addition, using nationwide data, the Insurance Information Institute (2010a, b) reports that the typical auto collision and auto comprehensive claims were around \$3,000 and \$1,500 in 2008 and the typical home claim exceeded \$8,000 in 2007.

somewhat (from 0.48 and 0.42 to 0.56 and 0.46), as does the threewise success rate (from 0.23 to 0.27), but each remains well below (and outside the confidence interval for) its corresponding expected success rate.

Moral Hazard.—We assume that a household's deductible choices do not affect its claim rates—i.e., there is no moral hazard. Most studies that test for the presence of asymmetric information in auto insurance markets using cross-sectional data do not find evidence of a positive correlation between risk and coverage (Pierre-André Chiappori 1999; Chiappori and Salanié 2000; Chiappori et al. 2006; Dionne, Gouriéroux, and Vanasse 1999, 2001).⁴⁵ These studies have been interpreted as casting doubt on the presence of moral hazard (Cohen 2005), at least in auto insurance. Recently, a handful of papers separately test for moral hazard in longitudinal auto insurance data using a dynamic approach pioneered by Japp H. Abbring, Chiappori, and Jean Pinquet (2003) and Abbring et al. (2003). Abbring, Chiappori, and Pinquet (2003) find no evidence of moral hazard in French auto insurance data. Mark Israel (2004) reports a small, but statistically significant, moral hazard effect for drivers in Illinois. Abbring, Chiappori, and Tibor Zavadil (2008), Olivia Ceccarini (2008), Pinquet et al. (2008), and Dionne, Pierre-Carl Michaud, and Maki Dahchour (2010) present stronger evidence of moral hazard using auto insurance data from Italy, France, the Netherlands, and Quebec, respectively. Each of the foregoing papers, 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 or quantifies the separate moral hazard effect, if any, directly attributable to the auto coverages we study, i.e., collision and comprehensive. Even if moral hazard were an issue with respect to any of the coverages we study,⁴⁶ it seems unlikely that claim rates would be significantly affected by the choice of deductible. Each of the deductible options typically represents a small fraction of the insured value of the household's auto or home. As a result, the choice of deductible has a small effect on the household's overall level of coverage.

State Independence.—The model assumes that the Bernoulli utility function is state independent. State dependent utility surely is warranted when modeling households' choices regarding flight insurance (e.g., Robert Eisner and Robert H. Strotz 1961; Edi Karni 1983), health and life insurance (e.g., Richard J. Zeckhauser 1970, 1973; Arrow 1974; Karni 1985; Karni and Itzhak Zilcha 1985; Viscusi and Evans 1990; Evans and Viscusi 1991), and insurance of irreplaceable commodities (Philip J. Cook and Daniel A. Graham 1977). When modeling (as we do) households' choices regarding insurance of replaceable property (autos and homes), however, state independent utility seems justified.

⁴⁵We are aware of two exceptions: Robert Puelz and Arthur Snow (1994) and Cohen (2005). Recent work criticizes the methodology of Puelz and Snow (1994) and suggests that their results are spurious (Chiappori 1999; Dionne, Gouriéroux, and Vanasse 1999). Cohen (2005) presents mixed evidence of asymmetric information. She finds that a positive coverage-accident correlation exists for new customers with more than three years of driving experience, but not for new customers with fewer than three years of driving experience. To our knowledge, there are no studies that test for asymmetric information in home insurance.

⁴⁶To the extent that a household's claim rates in auto comprehensive and home reflect force majeure risk more than household behavior risk, there is reason to believe that moral hazard is not a significant issue in the case of these coverages.

Independence of Claim Rates.—The regression model assumes that within-household claim rates are independent across coverages. This seems reasonable insofar as a household's claim rates in auto comprehensive and home reflect force majeure risk more than household behavior risk. As a check of this assumption, we compute pairwise correlations between the observed claim rates for households in the Test sample for whom we have claims information. There is essentially no correlation between the claim rates in auto and home and a weak positive correlation between the auto claim rates.⁴⁷ This suggests that, assuming within-household independence of claim rates is conservative, the patterns of empty intersections documented in Section IVC imply that a strong negative correlation between auto and home (particularly between auto comprehensive and home) would be required to markedly improve the success rates. Furthermore, insofar as there is no unobserved heterogeneity in auto collision, neither the pairwise success rate test nor the alternative test requires the independence assumption.

VI. Concluding Remarks

We examine the hypothesis of stable risk preferences using a unique dataset that documents the choices of a sample of households under their auto and home insurance policies. For each household, we test whether its deductible choices across three types of coverage—auto collision, auto comprehensive, and home all perils—can be rationalized by the same coefficient of absolute risk aversion. We find that the hypothesis of stable risk preferences is rejected by the data. This is not to say that there is no systematic component to the households' deductible choices. After all, the model is quite consistent with the households' single context choices (particularly under the assumption of CARA utility) and outperforms a fully random deductible choice model in terms of explaining the households' joint choices. Overall, however, our results call into question the empirical validity of the assumption of context-invariant risk preferences and caution against extrapolating estimates of risk preferences across contexts.

What can explain our results? Three alternatives immediately come to mind. First, households may have incorrect subjective beliefs about their claim rates. Although determining the subjective beliefs of the households in our data is beyond the scope of the present study (we cannot directly assess their beliefs, and estimating their beliefs from the data would not be straightforward), our analyses of two related explanations—unobserved heterogeneity and probability weighting—suggest that incorrect subjective beliefs may not be sufficient to explain the results. After all, while they offer different economic explanations for our results, each of unobserved heterogeneity, probability weighting, and incorrect subjective beliefs offers the same mathematical explanation: the claim rates that households use in their expected utility calculus do not equal the predicted claim rates that we use in our test.⁴⁸ More

⁴⁷ Neither the correlation between auto collision and home nor the correlation between auto comprehensive and home is statistically significant at the 10 percent confidence level. The correlation between auto collision and auto comprehensive is 0.12 and is significant at the 10 percent confidence level.

⁴⁸ The respective economic explanations run as follows: (i) in the case of unobserved heterogeneity, households know and use their objective claim rates in their expected utility calculus, but the predicted claim rates are incorrect; (ii) in the case of probability weighting, households know their objective claim rates and the predicted claim

specifically, our analysis of unobserved heterogeneity suggests that households' subjective beliefs would have to be far from the truth to plausibly explain the results. Moreover, patterns in the results suggest that households' subjective claim rates in auto comprehensive and home would have to be strongly negatively correlated, which seems counterintuitive insofar as such coverages primarily insure against events and circumstances beyond households' control. Our analysis of probability weighting suggests that systematically incorrect subjective beliefs cannot explain the results, at least to the extent that such beliefs mathematically correspond to the decision weights that result from conventional probability weighting.

The second alternative is that households are making mistakes. Such mistakes could be incorporated into the model by adding an error term to equation (1) (John D. Hey and Chris Orme 1994). While we leave the task of specifying a model of deductible choice with mistakes and estimating the distribution of the errors to future research, we can use our model to get an idea of the magnitude of the cost of such errors. Assuming quadratic utility, treating its predicted claim rate as its true claim rate, and taking the midpoint of its auto collision test interval as its true coefficient of absolute risk aversion, we calculate the error cost for each household who makes a mistake (i.e., who does not choose the optimal deductible given its true claim rate and coefficient of absolute risk aversion) as the difference between the certainty equivalent of the lottery induced by its optimal deductible choice and the certainty equivalent of the lottery induced by its actual deductible choice. For households who make mistakes (and have nonempty auto collision intervals), we find that the average error cost is \$5 in the case of auto comprehensive mistakes (or 6 percent of the mean auto comprehensive premium in the Test sample) and \$62 in the case of home mistakes (or 11 percent of the mean home premium in the Test sample).

The third alternative is that households have stochastic preferences. One model of stochastic preferences that appears promising at first glance is the random preference model (Graham Loomes and Robert Sugden 1995). Under this model, the decision maker has core preferences with one or more stochastic parameters. For example, were we to specify a random preference model of deductible choice, the core preferences would be represented by equation (1) and the stochastic parameter would be the household's coefficient of absolute risk aversion. A key feature of the random preference model is that it allows the variance of the stochastic parameter to decrease as the stakes of the choice increase. This feature appears to be consistent with our data. In addition to deductible choices, the data record choices of limits on liability coverages in auto (bodily injury and property damage) and home (personal liability). The stakes involved in the choice of liability limits are large relative to the stakes involved in the choice of deductibles—the differences among limit options constitute tens and hundreds of thousands of dollars. When we plot the distribution of personal liability limits conditional on the liability limit for bodily injury or property, there appears to be a strong correlation among liability limits across contexts, which suggests that households' liability limit choices reflect more stable risk preferences than do their deductible choices. Although this fits nicely with the story

rates are correct, but households use systematic distortions of their objective claim rates as decision weights in their expected utility calculus; and (iii) in the case of incorrect subjective beliefs, the predicted claim rates are correct but households neither know nor use their objective claim rates in their expected utility calculus.

of stochastic preferences having variance inversely related to stakes, confirmation via our test is outside the scope of this study (estimating liability claim rates would be difficult due to data limitations). Furthermore, we wonder whether the variance required to rationalize the deductible choices under a stochastic preferences model would strike us as reasonable; our analysis of unobserved heterogeneity suggests that it would be quite high.

Having previewed directions for future research, let us recap the contributions of the present study. The principal contribution is that we present direct evidence on the hypothesis that risk preferences are stable across multiple decision contexts using data on market choices. In doing so, we believe that we bridge a gap between the risk preferences literature, which provides estimates of risk preferences using unicontext market data, and the invariance literature, which documents violations of invariance using multicontext nonmarket data. A related contribution is that we assemble a new dataset that records multiple choices by the same economic units in market settings that are characterized by choice sets containing comparable money lotteries. Indeed, we presume that the aforementioned gap in the extant literature results from the paucity of data with these features being available for academic use. Lastly, we develop a method for testing the invariance of structural parameters of rational choice models. Notably, both our data and methods are highly portable—the data are well suited to addressing numerous other research questions concerning risk preferences, and our test is readily adaptable to other models of decision making under risk and other similar datasets.

APPENDIX

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).

B. Size and Power of the Test

The first step in calculating the size of the test is generating hypothetical data under the restriction that households have stable risk preferences. For each household i : (i) we draw a unique coefficient of absolute risk aversion from a uniform distribution over its benchmark auto collision interval; (ii) we take $\hat{\lambda}_{iL}$ as its auto collision claim rate; and (iii) we draw $\exp(\varepsilon_{iM})$ from $\text{gamma}(1/\hat{\alpha}_M, \hat{\alpha}_M)$ and $\exp(\varepsilon_{iH})$ from $\text{gamma}(1/\hat{\alpha}_H, \hat{\alpha}_H)$ and take $\hat{\lambda}_{iM} \times \exp(\varepsilon_{iM})$ and $\hat{\lambda}_{iH} \times \exp(\varepsilon_{iH})$ as its

auto comprehensive and home claim rates. We then determine each household's optimal deductible choices for auto collision, auto comprehensive, and home coverage given its unique coefficient of absolute risk aversion and its claim rates. The next step is performing the test on the hypothetical data—i.e., calculating the test intervals, the success rates, the expected success rates, and confidence intervals for the expected success rates. In the final step, we repeat the first two steps 1,000 times and tabulate the frequency with which we (incorrectly) reject the null hypothesis at standard levels of significance. We find that the size of the test is: 1.7 percent at the 2.5 percent level; 0.8 percent at the 1 percent level; 3.8 percent at the 5 percent level; and 8.9 percent at the 10 percent level.

The first step in calculating the power of the test is generating hypothetical data under the alternative hypothesis that households do not have stable risk preferences. For each household i , we take a draw from a uniform distribution over its benchmark auto collision interval and treat the draw as the household's coefficient of absolute risk aversion for auto collision, r_{iL} . We then construct the household's coefficients of absolute risk aversion for auto comprehensive and home as follows:

$$r_{iM} = (1 + \delta X_{iM}) r_{iL};$$

$$r_{iH} = (1 + \delta X_{iH}) r_{iL},$$

where X_{iM} and X_{iH} are independently drawn from $uniform(-1, 1)$ and $\delta > 0$ is a multiplier that controls the magnitude of the deviation of the alternative hypothesis from the null hypothesis $r_{iL} = r_{iM} = r_{iH}$.⁴⁹ Next, (i) we take $\hat{\lambda}_{iL}$ as the household's auto collision claim rate, and (ii) we draw $\exp(\varepsilon_{iM})$ from $gamma(1/\hat{\alpha}_M, \hat{\alpha}_M)$ and $\exp(\varepsilon_{iH})$ from $gamma(1/\hat{\alpha}_H, \hat{\alpha}_H)$ and take $\hat{\lambda}_{iM} \times \exp(\varepsilon_{iM})$ and $\hat{\lambda}_{iH} \times \exp(\varepsilon_{iH})$ as its auto comprehensive and home claim rates. We then determine each household's optimal deductible choices for auto collision, auto comprehensive, and home given its coefficients of absolute risk aversion and its claim rates.

Once we generate hypothetical data for a given multiplier δ , the second step is performing the test on the hypothetical data—i.e., calculating the test intervals, the success rates, the expected success rates, and confidence intervals for the expected success rates. In the final step, we repeat the first two steps 1,000 times and tabulate the frequency with which we (correctly) reject the null hypothesis at standard levels of significance. Table A1 reports the results for various multipliers between 0.25 and 1.25. The power behaves as expected at each significance level; i.e., the power is low for small values of δ and high for large values of δ .

C. Alternative Test Results

Tables A2 and A3 present the results of the alternative test for auto comprehensive coverage and home coverage, respectively.

In calculating the size of the alternative test, we first generate hypothetical data under the restriction that households have stable risk preferences using the same

⁴⁹To keep r_{iM} and r_{iH} within reasonable bounds, we impose lower and upper bounds equal to twice the minimum finite lower bound and twice the maximum finite upper bound of the relevant test intervals in the Test sample.

TABLE A1—POWER OF THE TEST

Significance level	Multiplier				
	0.25	0.50	0.75	1.00	1.25
1.0 percent	0.008	0.042	0.216	0.791	0.994
2.5 percent	0.024	0.085	0.329	0.895	0.996
5.0 percent	0.043	0.136	0.451	0.940	0.998
10.0 percent	0.090	0.222	0.589	0.967	1.000

TABLE A2—ALTERNATIVE TEST RESULTS: AUTO COMPREHENSIVE COVERAGE

Sample/robustness check	χ^2 critical value at 1% significance level	Minimum value of test statistic	1% value of test statistic	Rejection rate (percent)
<i>Test sample</i>				
Benchmark claim rates	11.34	211.00	221.77	100
Conditional claim rates	11.34	225.00	247.26	100
Test sample—2006 subsample	11.34	20.76	26.76	100
<i>Intersection subsample</i>				
Benchmark claim rates	11.34	41.28	41.28	100
Subsample claim rates	11.34	125.53	125.53	100
Conditional claim rates	11.34	86.61	103.82	100
Auto sample	11.34	689.11	720.72	100
Auto sample—2006 subsample	11.34	248.66	282.15	100
CARA utility	11.34	150.81	169.95	100
Probability weighting				
$\gamma = 0.16$	11.34	474.51	487.55	100
$\gamma = 0.40$	11.34	3,728.00	3,809.34	100
$\gamma = 0.69$	11.34	1,035.83	1,065.96	100
$\gamma = 0.85$	11.34	402.34	434.34	100
Prospect theory	11.34	796.76	855.82	100
Transaction costs (\$500)	11.34	211.00	221.77	100

TABLE A3—ALTERNATIVE TEST RESULTS: HOME COVERAGE

Sample/robustness check	χ^2 critical value at 1% significance level	Minimum value of test statistic	1% value of test statistic	Rejection rate (percent)
<i>Test sample</i>				
Benchmark claim rates	9.21	759.63	789.62	100
Conditional claim rates	9.21	683.85	740.09	100
Test sample—2006 subsample	9.21	5.34	10.74	99.5
<i>Intersection subsample</i>				
Benchmark claim rates	9.21	186.43	203.07	100
Subsample claim rates	9.21	66.93	81.89	100
Conditional claim rates	9.21	194.03	210.95	100
CARA utility	9.21	166.41	185.80	100
Probability weighting				
$\gamma = 0.16$	9.21	59.17	62.51	100
$\gamma = 0.40$	9.21	1,472.28	1,484.55	100
$\gamma = 0.69$	9.21	761.61	814.32	100
$\gamma = 0.85$	9.21	683.27	744.13	100
Prospect theory	9.21	1,534.40	1,601.30	100
Transaction costs (\$500)	9.21	395.68	462.20	100

TABLE A4—SIZE OF THE ALTERNATIVE TEST

Level implied by “rejection rate” rule	Level associated with Wald ellipsoid			
	1 percent	2.5 percent	5 percent	10 percent
1 percent	0.000	0.000	0.000	0.000
5 percent	0.000	0.000	0.000	0.000
10 percent	0.000	0.000	0.000	0.000
20 percent	0.000	0.000	0.000	0.010
50 percent	0.000	0.006	0.019	0.060
75 percent	0.016	0.052	0.101	0.237
90 percent	0.085	0.202	0.350	0.606
95 percent	0.194	0.373	0.594	0.852
99 percent	0.550	0.803	0.985	1.000

procedures described above. We then perform the alternative test on the hypothetical data—i.e., calculate ϕ_j , generate Φ_j 1,000 times, and verify how often Φ_j belongs to the Wald confidence ellipsoid for ϕ_j (see Section VA). Finally, we repeat the first two steps 1,000 times and tabulate the frequency with which we (incorrectly) reject the null hypothesis of stable risk preferences at various combinations of the test’s two levels, namely, (i) the level associated with the Wald confidence ellipsoid and (ii) the level implied by the rejection rule.⁵⁰ The results are set forth in Table A4.⁵¹ As we can see, the size of the alternative test is quite small at low levels. This is hardly surprising. After all, if households have stable risk preferences, then we would expect Φ_j to lie inside the Wald confidence ellipsoid of ϕ_j very frequently. Therefore, if we adopt a rejection rule pursuant to which we reject the null hypothesis of stable risk preferences only if Φ_j lies outside the Wald confidence ellipsoid very frequently, then the probability of making a Type I error should be quite low. Consider, for example, the level combination (1 percent, 1 percent). At the true \mathbf{r} and λ , the model-generated distribution would lie outside Wald confidence ellipsoid (and thereby lead us to incorrectly reject the null) approximately 1 percent of the time. Our approach makes the test even more conservative. First, the way we fix \mathbf{r} effectively minimizes the distance between the empirical and model-generated distributions for any given λ . Second, insofar as some realizations of λ could further reduce the gap between the data and the model, requiring that only 10 out of 1,000 realizations of λ generate a distribution Φ_j that lies inside the Wald confidence ellipsoid essentially reduces the probability of making a Type I error to nil.

In calculating the power of the alternative test, we first generate hypothetical data under the alternative hypothesis that households do not have stable risk preferences using the same procedures described above. Once we generate hypothetical data for a given multiplier δ , we then perform the alternative test on the hypothetical data—i.e., calculate ϕ_j , generate Φ_j 1,000 times, and verify how often Φ_j belongs to the Wald confidence ellipsoid for ϕ_j (see Section VA). Finally, we repeat the first two steps 1,000 times and tabulate the frequency with which we (correctly) reject

⁵⁰For instance, (i) if we use the 99 percent Wald confidence ellipsoid, the associated level is 1 percent, and (ii) if we reject the null hypothesis if Φ_j lies outside \mathcal{W}_j at least 990 times out of 1,000, the implied level is 1 percent.

⁵¹Note that the results of the alternative test reported in Section VA are based on the level combination (1 percent, 1 percent).

TABLE A5—POWER OF THE ALTERNATIVE TEST

Level combination	Multiplier				
	0.25	0.50	0.75	1.00	1.25
(1 percent, 1 percent)	0.002	0.001	0.006	0.137	0.890
(1 percent, 69.8 percent)	0.010	0.012	0.022	0.320	0.958
(5 percent, 42.7 percent)	0.012	0.015	0.045	0.408	0.975
(10 percent, 26.2 percent)	0.011	0.013	0.046	0.420	0.977
(20 percent, 12.7 percent)	0.008	0.014	0.049	0.451	0.979
(30 percent, 7.4 percent)	0.010	0.018	0.060	0.492	0.983

Notes: Size is 1 percent for each level combination, except level combination (1 percent, 1 percent) for which the size is less than 0.1 percent.

the null hypothesis of stable risk preferences at six combinations of the test's two levels. Table A5 reports the results for various multipliers between 0.25 and 1.25. The power behaves as expected at each test level combination.

REFERENCES

- Abbring, Jaap H., Pierre-André Chiappori, and Jean Pinquet.** 2003. "Moral Hazard and Dynamic Insurance Data." *Journal of the European Economic Association*, 1(4): 767–820.
- Abbring, Jaap H., Pierre-André Chiappori, and Tibor Zavadil.** 2008. "Better Safe than Sorry? Ex Ante and Ex Post Moral Hazard in Dynamic Insurance Data." Tilburg University Center for Economic Research Discussion Paper 2008-77.
- Abbring, Jaap H., Pierre-André Chiappori, James Heckman, and Jean Pinquet.** 2003. "Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?" *Journal of the European Economic Association*, 1(2–3): 512–21.
- Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström.** 2008. "Lost in State Space: Are Preferences Stable?" *International Economic Review*, 49(3): 1091–1112.
- Anderson, Lisa R., and Jennifer M. Mellor.** 2009. "Are Risk Preferences Stable? Comparing an Experimental Measure with a Validated Survey-Based Measure." *Journal of Risk and Uncertainty*, 39(2): 137–60.
- Arrow, Kenneth J.** 1951. "Alternative Approaches to the Theory of Choice in Risk-Taking Situations." *Econometrica*, 19(4): 404–37.
- Arrow, Kenneth J.** 1974. "Optimal Insurance and Generalized Deductibles." *Scandinavian Actuarial Journal*, 1: 1–42.
- Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro.** 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112(2): 537–79.
- Baucells, Manel, and Antonio Villasís.** 2010. "Stability of Risk Preferences and the Reflection Effect of Prospect Theory." *Theory and Decision*, 68(1–2): 193–211.
- Blavatsky, Pavlo, and Ganna Pogrebna.** 2008. "Risk Aversion When Gains Are Likely and Unlikely: Evidence from a Natural Experiment with Large Stakes." *Theory and Decision*, 64(2–3): 395–420.
- Boyer, Marcel, Georges Dionne, and Charles Vanasse.** 1992. "Econometric Models of Accident Distributions." In *Contributions to Insurance Economics*, ed. G. Dionne, 169–213. Norwell, MA: Kluwer Academic.
- Camerer, Colin.** 1995. "Individual Decision Making." In *The Handbook of Experimental Economics*, ed. J. H. Kagel and A. E. Roth, 587–703. Princeton, NJ: Princeton University Press.
- Cameron, A. Colin, and Pravin K. Trivedi.** 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics*, 1(1): 29–53.
- Cameron, A. Colin, and Pravin K. Trivedi.** 1998. *Regression Analysis of Count Data*. Econometric Society Monograph 30. Cambridge: Cambridge University Press.
- Ceccarini, Olivia.** 2008. "Does Experience Rating Matter in Reducing Accident Probabilities? A Test for Moral Hazard." PhD diss. University of Pennsylvania.

- Chetty, Raj.** 2006. "A New Method of Estimating Risk Aversion." *American Economic Review*, 96(5): 1821–34.
- Chiappori, Pierre-André.** 1999. "Asymmetric Information in Automobile Insurance: An Overview." In *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, ed. G. Dionne and C. Laberge-Nadeau, 1–11. Norwell, MA: Kluwer Academic.
- Chiappori, Pierre-André, and Bernard Salanié.** 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108(1): 56–78.
- Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and Francois Salanié.** 2006. "Asymmetric Information in Insurance: General Testable Implications." *RAND Journal of Economics*, 37(4): 783–98.
- Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv.** 2007. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, 97(5): 1921–38.
- Cicchetti, Charles J., and Jeffrey A. Dubin.** 1994. "A Microeconomic Analysis of Risk Aversion and the Decision to Self-Insure." *Journal of Political Economy*, 102(1): 169–86.
- Cohen, Alma.** 2005. "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market." *Review of Economics and Statistics*, 87(2): 197–207.
- Cohen, Alma, and Liran Einav.** 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review*, 97(3): 745–88.
- Cook, Philip J., and Daniel A. Graham.** 1977. "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *Quarterly Journal of Economics*, 91(1): 143–56.
- Dionne, Georges, Christian Gouriéroux, and Charles Vanasse.** 1999. "Evidence of Adverse Selection in Automobile Insurance Markets." In *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, ed. G. Dionne and C. Laberge-Nadeau, 13–46. Norwell, MA: Kluwer Academic.
- Dionne, Georges, Christian Gouriéroux, and Charles Vanasse.** 2001. "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment." *Journal of Political Economy*, 109(2): 444–53.
- Dionne, Georges, Pierre-Carl Michaud, and Maki Dahchour.** 2010. "Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France." Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi Working Paper 10–35.
- Dionne, Georges, and Charles Vanasse.** 1992. "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information." *Journal of Applied Econometrics*, 7(2): 149–65.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner.** Forthcoming. "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences." *Journal of the European Economic Association*.
- Economides, Nicholas.** 1982. "The Demand for Life Insurance: An Application of the Economics of Uncertainty: A Comment." *Journal of Finance*, 37(5): 1305–09.
- Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen.** 2010. "How General Are Risk Preferences? Choices under Uncertainty in Different Domains." National Bureau of Economic Research Working Paper 15686.
- Eisner, Robert, and Robert H. Strotz.** 1961. "Flight Insurance and the Theory of Choice." *Journal of Political Economy*, 69(4): 355–368.
- Evans, William N., and W. Kip Viscusi.** 1991. "Estimation of State-Dependent Utility Functions Using Survey Data." *Review of Economics and Statistics*, 73(1): 94–104.
- Gouriéroux, Christian, Alain Monfort, and Alain Trognon.** 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica*, 52(3): 701–20.
- Greenwood, Major, and G. Udny Yule.** 1920. "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents." *Journal of the Royal Statistical Society*, 83(2): 255–79.
- Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches.** 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, 52(4): 909–38.
- Hershey, John C., and Paul J. H. Schoemaker.** 1980. "Risk Taking and Problem Context in the Domain of Losses: An Expected Utility Analysis." *Journal of Risk and Insurance*, 47(1): 111–32.
- Hershey, John C., Howard C. Kunreuther, and Paul J. H. Schoemaker.** 1982. "Sources of Bias in Assessment Procedures for Utility Functions." *Management Science*, 28(8): 936–54.
- Herzog, Sergio.** 2002. "Empirical Analysis of Motor Vehicle Theft in Israel, 1990–1997." *British Journal of Criminology*, 42(4): 709–28.
- Hey, John D., and Chris Orme.** 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica*, 62(6): 1291–1326.

- Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5): 1644–55.
- Insurance Information Institute. 2010a. "Auto Insurance." <http://www.iii.org/media/facts/statsbyissue/auto/> (accessed January 11, 2010).
- Insurance Information Institute. 2010b. "Homeowners Insurance." <http://www.iii.org/media/facts/statsbyissue/homeowners/> (accessed January 11, 2010).
- International Traffic Safety Data and Analysis Group. 2009. "IRTAD Annual Report 2009." <http://internationaltransportforum.org/irtad/pdf/09IrtadReportFin.pdf>.
- Israel, Mark. 2004. "Do We Drive More Safely When Accidents are More Expensive? Identifying Moral Hazard from Experience Rating Schemes." Northwestern University Center for the Study of Industrial Organization Working Paper 0043.
- Johnson, Eric J., and Daniel G. Goldstein. 2003. "Do Defaults Save Lives?" *Science*, 302(5649): 1338–39.
- Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther. 1993. "Framing, Probability Distortions, and Insurance Decisions." *Journal of Risk and Uncertainty*, 7(1): 35–51.
- Jullien, Bruno, and Bernard Salanié. 2000. "Estimating Preferences under Risk: The Case of Race-track Bettors." *Journal of Political Economy*, 108(3): 503–30.
- Kachelmeier, Steven J., and Mohamed Shehata. 1992. "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China." *American Economic Review*, 82(5): 1120–41.
- Kahneman, Daniel. 2003a. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5): 1449–75.
- Kahneman, Daniel. 2003b. "A Psychological Perspective on Economics." *American Economic Review*, 93(2): 162–68.
- Karni, Edi. 1983. "Risk Aversion for State-Dependent Utility Functions: Measurement and Applications." *International Economic Review*, 24(3): 637–47.
- Karni, Edi. 1985. *Decision Making Under Uncertainty: The Case of State-Dependent Preferences*. Cambridge, MA: Harvard University Press.
- Karni, Edi, and Itzhak Zilcha. 1985. "Uncertain Lifetime, Risk Aversion, and Life Insurance." *Scandinavian Actuarial Journal*, 1985: 109–23.
- Kimball, Miles S. 1990. "Precautionary Saving in the Small and in the Large." *Econometrica*, 58(1): 53–73.
- Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2008. "Imputing Risk Tolerance from Survey Responses." *Journal of the American Statistical Association*, 103(483): 1028–38.
- Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2009. "Risk Preferences in the PSID: Individual Imputations and Family Covariation." *American Economic Review*, 99(2): 363–68.
- Krantz, David H., and Howard C. Kunreuther. 2007. "Goals and Plans in Decision Making." *Judgment and Decision Making*, 2(3): 137–68.
- Loomes, Graham, and Robert Sugden. 1995. "Incorporating a Stochastic Element into Decision Theories." *European Economic Review*, 39(3–4): 641–48.
- McCaffery, Edward J. 1994. "Cognitive Theory and Tax." *UCLA Law Review*, 41(7): 1861–47.
- Novemsky, Nathan, and Daniel Kahneman. 2005. "The Boundaries of Loss Aversion." *Journal of Marketing Research*, 42(2): 119–28.
- Pinquet, Jean, Georges Dionne, Charles Vanasse, and Mathieu Maurice. 2008. "Point-Record Incentives, Assymmetric Information and Dynamic Data." Centre National de la Recherche Scientifique Working Paper 2008–05.
- Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1): 38–71.
- Pratt, John W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica*, 32(1–2): 122–36.
- Puelz, Robert, and Arthur Snow. 1994. "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market." *Journal of Political Economy*, 102(2): 236–57.
- Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature*, 36(1): 11–46.
- Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.
- Rabin, Matthew, and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives*, 15(1): 219–32.
- Saha, Atanu. 1997. "Risk Preference Estimation in the Nonlinear Mean Standard Deviation Approach." *Economic Inquiry*, 35(4): 770–82.
- Samuelson, Paul A. 1963. "Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia*, 98: 108–13.
- Schelling, Thomas C. 1981. "Economic Reasoning and the Ethics of Policy." *Public Interest*, 63: 37–61.

- Schoemaker, Paul J. H., and Howard C. Kunreuther.** 1979. "An Experimental Study of Insurance Decisions." *Journal of Risk and Insurance*, 46(4): 603–18.
- Starmer, Chris.** 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature*, 38(2): 332–82.
- Sydnor, Justin.** 2010. "(over)Insuring Modest Risks." *American Economic Journal: Applied Economics*, 2(4): 177–99.
- Sydnor, Justin.** 2006. "Abundant Aversion to Moderate Risk: Evidence from Homeowners Insurance." http://faculty.weatherhead.case.edu/sydnor/deductibles_old.pdf.
- Thaler, Richard H.** 1980. "Towards a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1(1): 39–60.
- Tversky, Amos, and Daniel Kahneman.** 1986. "Rational Choice and the Framing of Decisions." *Journal of Business*, 59(4): S251–78.
- Tversky, Amos, and Daniel Kahneman.** 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.
- Tversky, Amos, and Richard H. Thaler.** 1990. "Anomalies: Preference Reversals." *Journal of Economic Perspectives*, 4(2): 201–11.
- Viscusi, W. Kip, and William N. Evans.** 1990. "Utility Functions That Depend on Health Status: Estimates and Economic Implications." *American Economic Review*, 80(3): 353–74.
- Wolf, Charles, and Larry Pohlman.** 1983. "The Recovery of Risk Preferences from Actual Choices." *Econometrica*, 51(3): 843–50.
- Wooldridge, Jeffrey M.** 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Zeckhauser, Richard.** 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory*, 2(1): 10–26.
- Zeckhauser, Richard J.** 1973. "Coverage for Catastrophic Illness." *Public Policy*, 21(2): 149–72.