

Different Contexts, Different Risk Preferences?

Levon Barseghyan* Joshua C. Teitelbaum† Lin Xu‡

Draft: September 2, 2022

Abstract

We examine the stability of risk preferences across contexts involving different stakes. Using data on households' deductible choices in three property insurance coverages and their limit choices in two liability insurance coverages, we assess the stability across the five contexts in the ordinal ranking of the households' willingness to bear risk. We find evidence of stability across contexts involving stakes of the same magnitude, but not across contexts involving stakes of very different magnitudes. We explore potential explanations for these patterns including heterogeneity in wealth, access to credit, and limited consideration. (*JEL* D12, D81, G22)

*Cornell University, Department of Economics, lb247@cornell.edu.

†Georgetown University, Law Center and Department of Economics, jct48@georgetown.edu.

‡United States Congress, Joint Committee on Taxation, lx82@cornell.edu.

1 Introduction

Classical theories of risky choice posit that risk preferences are stable across decision contexts. The stability hypothesis reflects a basic tenet of rational choice theory known as invariance (Tversky and Kahneman 1986) or context independence (Hausman 2012). Context independence requires that preferences over options be invariant to the aspects of the choice situation other than the economic fundamentals, which in the case of risky options are the induced lotteries over outcomes.

Broadly speaking, the empirical literature on the stability hypothesis offers two main findings. On the one hand, studies that focus on the (strong) hypothesis of full stability—which usually take a structural approach and examine the within-person consistency of model-based estimates of risk aversion across domains—generally find that a person’s risk aversion differs from one domain to the next, suggesting that risk preferences are not perfectly stable across contexts (e.g., Barseghyan, Prince, and Teitelbaum 2011). On the other hand, studies that focus on the (weak) hypothesis of some stability—which usually take a model-free approach and examine the within-person correlation of risk taking across domains—generally find that a person who takes on more risk in one context tends to do so in other contexts as well, suggesting that risk preferences have a stable component and are not entirely context dependent (e.g., Einav et al. 2012).

We provide new evidence on the stability hypothesis using data on households’ coverage choices in five insurance contexts. A key feature of our data is that three contexts involve small-stakes choices while two involve large-stakes choices. The small-stakes choices are deductibles in three lines of property insurance: auto collision, auto comprehensive, and home all perils. The large-stakes choices are limits in two lines of liability insurance: auto single limit and home personal liability. We adopt the model-free approach of Einav et al. (2012) and assess the stability in ranking across the five contexts of each household’s willingness to bear risk relative to its peers. Essentially, we rank the coverage options by

risk within each context and compute the pairwise rank correlations among the households' choices across the five contexts. In our preferred baseline specification, we estimate the rank correlations controlling for variation across households in the price of coverage and the risk of loss in each context.

Consistent with prior results in the literature, we find that the households' small-stakes choices are positively rank correlated. We also find that their large-stakes choices are positively rank correlated. Strikingly, however, we find that the households' small-stakes choices are negatively rank correlated with their large-stakes choices. That is, we find that a household who takes on more risk than its peers in small-stakes contexts tends to take on less risk than its peers in large-stakes contexts, and vice versa, which does not support the stability hypothesis, even in its weak form. We explore potential explanations for these patterns that would reconcile them with the stability hypothesis, including heterogeneity in wealth, access to credit, and limited consideration.

2 Related Literature

There are several previous empirical investigations of the stability hypothesis. We highlight a few key studies in the economics literature,¹ giving separate treatment to studies that use data on market choices and those that rely on data from experiments and surveys.

2.1 Studies Using Market Data

In an early paper, Wolf and Pohlman (1983) compare the risk aversion of a dealer in U.S. government securities as first measured by his assessments of hypothetical wealth gambles and then estimated from his bid choices in Treasury bill auctions. The authors

¹Barseghyan, Prince, and Teitelbaum (2011) and Galizzi, Miraldo, and Stavropoulou (2016) discuss additional studies. A separate literature investigates the consistency of risk preference measures obtained from different elicitation methods employed in experiments and surveys. For a summary, see Galizzi, Machado, and Miniaci (2016).

take a structural approach and assume the dealer is an expected utility (EU) maximizer. They find that "the dealer was substantially more risk averse in his bid choices than his assessments predicted" and conclude that people's "degree of risk aversion may depend on the specific context in which their choices are made" (p. 849).

Though pioneering, Wolf and Pohlman (1983) has two important limitations. First, it studies one person. Second, it compares hypothetical choices with market choices, which confounds the question of stability with that of external validity. Overcoming these limitations, Barseghyan, Prince, and Teitelbaum (2011) examine the deductible choices of 702 households across three insurance contexts: auto collision, auto comprehensive, and home all perils. Assuming that households are EU maximizers, the authors obtain three interval estimates of each household's risk aversion based on its three choices. They find that these intervals intersect—implying that the choices can be rationalized by the same degree of risk aversion—for only 23 percent of households, leading them to reject the hypothesis of full stability.

Rejecting the hypothesis of full stability does not imply that risk preferences have no stable component. Moreover, structural approaches to testing stability invariably comprise a joint test of the stability hypothesis and the assumptions of the structural model. With these points in mind, Einav et al. (2012) examine the workplace benefits choices made by 12,752 Alcoa employees in six contexts: health insurance, drug insurance, dental insurance, short-term disability insurance, long-term disability insurance, and 401(k) investments. The authors pursue a model-free approach (which we adopt here) in which they rank by risk the options within each context and assess the rank correlation of the employees' choices across the six contexts. They find that an employee's choice in each context is positively rank correlated with her choice in every other context, with stronger correlations

across "closer" contexts (p. 2609), leading them to reject the hypothesis of no stability and conclude that risk preferences have a context-invariant component.²

In the wake of this methodological shift, Barseghyan, Molinari, and Teitelbaum (2016) explore the connection between full stability under a structural approach and rank stability under a model-free approach. Using data on the deductible choices of 3,629 households across the three insurance contexts studied by Barseghyan, Prince, and Teitelbaum (2011), the authors document two findings: (i) the households' deductible choices are positively rank correlated, echoing the finding of Einav et al. (2012), and (ii) five in six households exhibit full stability under a rank-dependent EU model. They then show that the fully stable households drive the rank correlations.

Our paper builds directly on Einav et al. (2012). Like them, we take a model-free approach and examine rank stability across multiple contexts using data on market choices. The main distinction between our papers is the degree to which the stakes vary across contexts. In the contexts we study, the dollar values of the options range from the hundreds and thousands (in our small-stakes contexts) to the hundreds of thousands and millions (in our large-stakes contexts). As we discuss in Section 4.3, the dollar values of the options in Einav et al. (2012) range from the hundreds and thousands (in three contexts) to the tens of thousands (in the others). It is this distinction that reconciles our results. Both papers find evidence of rank stability across contexts involving stakes of the same or near orders of magnitude, while ours also finds evidence of rank instability across contexts involving stakes of remote orders of magnitude.³

²Einav et al. (2012) also pursue a structural approach that is conceptually similar to the approach in Barseghyan, Prince, and Teitelbaum (2011). Under this approach, they find that for roughly 30 percent of employees all six choices can be rationalized by the same degree of risk aversion.

³Collier et al. (2018) also study choices with remote stakes. Using data on households' deductibles and coverage limits in flood insurance, the authors structurally estimate the risk preferences implied by the two choices and find they differ. Because they take a structural approach, their paper relies on stronger modeling assumptions than ours. Indeed, their estimation approach—parametric MLE of a random utility model—entails even stronger assumptions than the partial identification approach taken by Barseghyan, Prince, and Teitelbaum (2011). In addition, the two choices they study are made in the same context. We therefore view their paper as more in line with the literature on how risk aversion varies with stake size (e.g., Binswanger 1980; Kachelmeier and Shehata 1992; Holt and Laury 2002; Fehr-Duda et al. 2010).

2.2 Studies Using Nonmarket Data

Anderson and Mellor (2009) compare the responses of laboratory subjects to a series of hypothetical job gambles and a series of hypothetical inheritance gambles. The authors construct a categorical measure of the subjects' risk aversion based on the job gamble responses and then do the same for the inheritance gamble responses. They find that 34 percent of subjects exhibit the same degree of risk aversion across the two contexts and report a rank correlation of 0.175 between the two measures.⁴ Dohmen et al. (2011) use survey data from the German Socio-Economic Panel to compare respondents' self-reported willingness to take risks across five contexts: car driving, financial matters, sports and leisure, health, and career. The authors report that while the responses "are not perfectly correlated across contexts, ... the pairwise correlations are large, typically in the neighborhood of 0.5," which they argue "is suggestive of a stable, underlying risk trait" (p. 537). Ioannou and Sadeh (2016) compare the selections made by laboratory subjects from a set of real monetary gambles and a set of real "environmental" gambles (where the payoffs are numbers of bee-friendly plants). The authors find that subjects "exhibit a higher degree of risk aversion in the environmental domain relative to the monetary domain; that is, individuals tend to be more reluctant to take on large gambles with environmental outcomes than with monetary ones" (p. 31).⁵ Charness et al. (2020) report the rank correlations of six measures of self-reported risk exposure (relating to investment, insurance, and employment choices) in a Dutch sample. The correlations are

⁴The job and inheritance gamble questions are taken from the Health and Retirement Study (HRS). Barksy et al. (1997) use the responses to the job gamble questions in the HRS to construct a measure of respondents' risk tolerance. They then present evidence that their measure predicts certain self-reported risky behaviors, "including smoking, drinking, not having insurance, choosing risky employment, and holding risky assets" (p. 551).

⁵In another incentivized experiment, Choi et al. (2007) test within-subject consistency across 50 risky portfolio choices. They find that while only 17 percent of subjects exhibit perfect consistency (app. C), a "significant majority" perform "only a bit worse" (pp. 1927-1928). More to the point, the authors report that "some subjects" exhibit a "switching" pattern—sometimes choosing extremely safe portfolios, sometimes choosing extremely risky portfolios, and sometimes choosing intermediate portfolios—wherein their choices are "individually consistent" with risk averse utility maximization but "mutually inconsistent" with one another (pp. 1925 & 1936-1937).

all weakly positive (the highest is 0.17) or statistically insignificant. They also test whether laboratory measures of risk attitude correlate with these risk-exposure measures or with hypothetical financial decisions (portfolio, insurance, mortgage). They find that most of the risk-attitude measures correlate with subjects' laboratory financial decisions, but that none correlate with their self-reported risk exposure. Jaspersen, Ragin, and Sydnor (2022) conduct an incentivized laboratory study in which subjects make both insurance choices and neutrally-framed money lottery choices. The authors use the lottery choices to measure risk-preference components (utility curvature, probability weighting, loss aversion) and report modest correlations between these measures and subjects' insurance choices.

In addition to using nonmarket data, these studies differ from ours in that they either lack meaningful variation in stakes across contexts (e.g., Anderson and Mellor 2009; Ioannou and Sadeh 2016; Jaspersen, Ragin, and Sydnor 2022) or they study domains of risky behavior in which the stakes are neither explicit nor well-defined (e.g., Barksy et al. 1997; Dohmen et al. 2011; Charness et al. 2020).

3 Data and Sample

The source of our data is a large U.S. property and casualty insurance company. Our dataset contains annual information on more than 400,000 households who purchased auto or home insurance from the company between 1998 and 2007. The data contain all the information in the company's records regarding the households and their policies, including claims information.

We focus on three small-stakes choices and two large-stakes choices. The small-stakes choices are deductibles in three lines of property coverage: auto collision, auto comprehensive, and home all perils. 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, without

regard to fault. Home all perils coverage pays for damage to the insured home from all causes, except those that are specifically excluded (e.g., flood). The deductible options range from \$100 to \$1,000 in auto collision, \$50 to \$1,000 in auto comprehensive, and \$100 to \$5,000 in home all perils. The mean increment between options is \$225 in auto collision, \$190 in auto comprehensive, and \$980 in home all perils.

The large-stakes choices are limits in two lines of liability coverage: auto single limit and home personal liability. Auto single limit coverage pays for bodily injury or property damage to others for which the insured driver is legally responsible. Home personal liability coverage pays for bodily injury or property damage to others for which the insured homeowner is legally responsible. The limit options range from \$60,000 to \$1,000,000 in auto single limit and \$100,000 to \$1,000,000 in home personal liability. The mean increment between options is \$188,000 in auto single limit and \$180,000 in home personal liability.

Our baseline sample comprises households who (i) purchased all three property coverages and both liability coverages and (ii) first purchased each coverage within any six-month window during the period from 2004 to 2007. The latter restriction helps avoid temporal issues, such as changes in household characteristics or the economic environment. We consider only the households' coverage choices at the time of first purchase. This helps ensure that we are working with active choices; one might worry that households renew their policies without actively reassessing their coverage options (Handel 2013). These restrictions yield a baseline sample of 2,690 households.

For each household in our baseline sample, we observe its deductible or limit choice (as the case may be) in each coverage, as well as the pricing menu it faced in each coverage. According to the company and an independent agent who sells company policies, the choice environment is conducive to households making active and informed choices—there are no default choices, the pricing menu is available to a household when it makes a choice, and a household must choose a deductible or limit separately for each coverage.

Tables 1 and 2 provide descriptive statistics for the baseline sample. Table 1 reports demographic characteristics and claim frequencies. Table 2 summarizes the coverage choices and pricing menus.

4 Methods and Results

4.1 Empirical Strategy and Baseline Results

We adopt the model-free approach of Einav et al. (2012) and assess the stability in ranking across contexts of each household’s willingness to bear risk relative to its peers. To begin, we rank the options by risk within each context, ordering them from highest to lowest risk exposure. There are five or six options in each context (see Table 2). The safest option is the lowest deductible in property coverages and the highest limit in the liability coverages. We then compute the pairwise Spearman rank correlations in the households’ choices across the five contexts.

Because these rank correlations do not control for potentially important covariates, we also examine the correlation structure of the residuals from a system of five equations:

$$\begin{bmatrix} y_i^{Auto\ collision} \\ y_i^{Auto\ comprehensive} \\ y_i^{Home\ all\ perils} \\ y_i^{Auto\ single\ limit} \\ y_i^{Home\ personal\ liability} \end{bmatrix} = \begin{bmatrix} \beta^{Auto\ collision} \\ \beta^{Auto\ comprehensive} \\ \beta^{Home\ all\ perils} \\ \beta^{Auto\ single\ limit} \\ \beta^{Home\ personal\ liability} \end{bmatrix} \cdot \mathbf{x}_i + \begin{bmatrix} \varepsilon_i^{Auto\ collision} \\ \varepsilon_i^{Auto\ comprehensive} \\ \varepsilon_i^{Home\ all\ perils} \\ \varepsilon_i^{Auto\ single\ limit} \\ \varepsilon_i^{Home\ personal\ liability} \end{bmatrix}, \quad (1)$$

where y_i^j denotes the rank-ordered choice of household i in context j , β^j is a vector of context-specific coefficients, \mathbf{x}_i is a vector of household-specific covariates, and ε_i^j is a household- and context-specific residual. In theory, a household’s choices depend not only on its risk preferences but also on the prices it faces and its risk profile. The baseline set of

covariates (\mathbf{x}_i), therefore, includes controls for prices and risk. The price controls are log-transformed premiums for each coverage assuming a \$250 deductible or \$200,000 limit, as the case may be.⁶ The risk controls are expected annual claims under each coverage based on separate Poisson-gamma Bayesian credibility models. By construction, the risk controls take into account both the systematic and idiosyncratic components of a households' risk type. For further details, see the Online Appendix.

Following Einav et al. (2012), we estimate system (1) in two different ways. First, we treat it as a multivariate ordered probit regression model and estimate it by maximum likelihood.⁷ Second, we treat it as a multivariate linear regression model and estimate it by least squares. Because the set of options in each context is discrete, the probit regression is our preferred specification.

Table 3 reports the baseline results. Panel A shows the Spearman rank correlations. Panels B and C display the estimated correlations from the probit and linear regressions, respectively. Each panel tells the same story. Across all panels, the correlation between each pair of small-stakes choices is positive, ranging from 0.26 to 0.70. Similarly, the correlation between the two large-stakes choices is positive, ranging from 0.44 to 0.57. By contrast, however, the correlation between every pairing of a small-stakes choice and a large-stakes choice is negative, ranging from -0.05 to -0.34 . Overall, the baseline results suggest that the households exhibit a fairly stable degree of risk aversion relative to their peers across contexts that involve stakes of the same order of magnitude. At the same time, however, the results suggest that households who exhibit a higher degree of risk aversion than their peers in small-stakes contexts tend to exhibit a lower degree of risk aversion than their peers in large-stakes contexts, and vice versa.

⁶We do not include a price control for home personal liability because the premiums do not vary across households.

⁷We estimate the system by performing bivariate ordered probit regressions on every pair of equations. In each regression, we use the Huber/White/sandwich estimator to obtain robust standard errors.

4.2 Sensitivity Analysis

4.2.1 Umbrella Coverage

Twenty-six percent of the households in the baseline sample purchased umbrella liability coverage from the company to supplement their auto single limit and home personal liability coverages. The umbrella coverage options range from \$1 million to \$5 million in \$1 million increments, and the premium associated with each coverage option is the same for all households.

The baseline results disregard the households' umbrella choices. To explore whether this biases our results, we treat households who purchased umbrella coverage as having chosen a new "highest limit" option (i.e., a limit of unspecified amount greater than \$1,000,000) in auto single limit and home personal liability, and we re-estimate (1) treating it as a system of ordered probits and including the baseline set of controls.⁸

Table 4, panel A reports the results, which tell the same story as the baseline results. Indeed, all but one of the pairwise correlations involving a liability insurance context are stronger than the corresponding baseline correlations. The only exception is the correlation between home all perils and home personal liability, which is slightly weaker than the corresponding baseline correlation.⁹

4.2.2 Wealth

Economists have long hypothesized that risk preferences depend on wealth (Friedman and Savage 1948; Pratt 1964; Arrow 1971). The standard assumption is that absolute risk aversion is decreasing in wealth, which implies that, *ceteris paribus*, a household's willingness to pay for insurance decreases with its wealth. See, e.g., Pratt (1964, pp. 122-

⁸We do not add a price control for umbrella coverage because the premiums do not vary across households.

⁹As a further check, we also re-estimate (1), again treating it as a system of ordered probits and including the baseline set of controls, on the subsample of 1,993 households who did not purchase umbrella coverage. Those results also tell the same story as the baseline results.

123): "Utility functions for which [the coefficient of absolute risk aversion] is decreasing are logical candidates to use when trying to describe the behavior of people who, one feels, might generally pay less for insurance against a given risk the greater their assets."

Our baseline analysis does not control for household wealth. To examine whether wealth effects may be driving our results, we add a control for wealth to the baseline set of controls and re-estimate (1) treating it as a system of ordered probits. We do not directly observe a household's wealth in our data, but we do observe a plausible proxy: the insured value of the dwelling covered by its homeowners policy ("home value"). Of course, we do not know the correlation between home value and wealth in our data. However, according to combined extract data (1989-2016) from the Survey of Consumer Finance, the correlation between home value and wealth is 0.47 (std. err. = 0.002).

Table 4, panel B reports the results. Each pairwise correlation is virtually identical to the corresponding baseline correlation. It thus appears that wealth effects are not driving our results.

4.2.3 Access to Credit

In theory, a household's ability to borrow after a loss event can affect its demand for insurance (Handel, Hendel, and Whinston 2015; Jaffe and Malani 2018). To investigate whether differences in access to credit may be driving our results, we add controls for households' insurance scores in auto and home to the baseline set of controls and re-estimate (1) treating it as a system of ordered probits. Insurance scores are akin to credit scores. Both are derived using the same five categories of information contained in credit reports (payment history, level of indebtedness, length of credit history, new credit and pursuit of new credit, and types of credit), though they differ somewhat in how they weight these categories (Morris, Schwarcz, and Teitelbaum 2017). For this reason, we believe that insurance score, like credit score, is a good proxy for a household's access to credit.¹⁰

¹⁰There is ample evidence that credit score is a good proxy for access to credit (e.g., Baker 2018).

Table 4, panel C reports the results. Again, each pairwise correlation is virtually identical to the corresponding baseline correlation. This suggests that differences in access to credit are not driving our results.¹¹

4.2.4 Choice Window

In the baseline sample, we restrict attention to households who, *inter alia*, purchased all five coverages within a six-month window. There are two opposing considerations in selecting a choice window. On the one hand, a narrower window helps to avoid what Einav et al. (2012, p. 2611) call "the problems of inferring preferences from 'stale' choices," which they note "could be particularly concerning if individuals might have made their choices . . . at different points in time." On the other hand, a wider window helps to improve inference by increasing sample size.

We are not concerned that a six-month window is too narrow. Our baseline sample comprises 2,690 households, which we believe is sufficiently large to draw valid inferences. To address the concern that a six-month window may be too wide, we re-estimate (1) on the subsample of 1,694 households who purchased all five coverages on the same day. As before, we treat (1) as a system of ordered probits and include the baseline set of controls. Table 4, panel D reports the results. They tell the same story as the baseline results. Indeed, all but two of the pairwise correlations are stronger than the corresponding baseline correlations. The only exceptions are the pairwise correlations between auto collision and home personal liability and between home all perils and home personal liability, which are slightly weaker than baseline.

¹¹As a further check, we re-estimate (1), again treating it as a system of ordered probits, with controls for wealth and insurance scores (and their interactions) added to the baseline set. Once again, the results tell the same story.

4.3 Comparison with Einav et al. (2012)

We close this section with a discussion comparing our results with those of Einav et al. (2012). Using data on the workplace benefits choices of 12,752 Alcoa employees, Einav et al. (2012) pursue the same model-free approach to assess the rank stability of the employees' risk preferences across six contexts: health insurance, drug insurance, dental insurance, short-term disability insurance, long-term disability insurance, and 401(k) investments. In their baseline analysis, where they control for variation in benefit menus, they find that an employee's choice in every context is positively rank correlated with its choice in every other context, implying that employees who exhibit a higher degree of risk aversion than their peers in one context tend also to do so in other contexts, and vice versa. They find very similar results when they add controls for risk. The strongest pairwise correlations are between short- and long-term disability insurance (0.76) and among health, drug, and dental insurance (ranging from 0.30 to 0.49). Somewhat weaker are the correlations across the disability and medical insurance contexts (ranging from 0.21 to 0.26). The weakest are between 401(k) investments and every other context (all below 0.05, including two that are slightly negative but not statistically different from zero).¹²

In order to compare our results with those of Einav et al. (2012), we must classify their contexts according to the magnitude of the stakes involved, applying the same criteria that we use to classify our contexts. Recall that in our small-stakes contexts the values of the options and the inter-option increments range in the hundreds and thousands dollars, whereas in our large-stakes contexts the value of the options range in the hundreds of thousands and millions of dollars with inter-option increments that range in the hundreds of thousands dollars.

For the reasons we detail in the Online Appendix, we conclude that none of the contexts in Einav et al. (2012) involve large-stakes choices. Specifically, we conclude that

¹²The quoted results are from Table 3B, panel A in Einav et al. (2012), which reports correlation estimates from a system of ordered probits with controls for benefit menus and risk.

three contexts—health, drug, and dental insurance—involve small-stakes choices. In two contexts—short-term disability insurance and 401(k) investments—we determine that the stakes range in the thousands and tens of thousands of dollars but not the hundreds of thousands of dollars, and so we classify them as moderate-stakes contexts. We also classify the remaining context—long-term disability insurance—as involving moderate-stakes choices, though the reasons are less straightforward.¹³

Given these classifications, we see that our results and those of Einav et al. (2012) complement one another. We both find a pattern of positive pairwise correlations among small-stakes choices. To this common result, Einav et al. (2012) add two findings: patterns of positive (or at least non-negative) pairwise correlations among moderate-stake choices and across small- and moderate-stakes choices. We also add two findings. The first is a pattern of positive pairwise correlations among large-stakes choices, which taken together with the previous findings hints at a stable component of risk preferences that operates across contexts involving stakes of the same or near orders of magnitude. The second finding that we add to the mix is our main contribution: a pattern of negative pairwise correlations across small- and large-stakes choices, which hints at a lack of risk preference stability across contexts involving stakes of remote orders of magnitude.

There is another way to see how we build on Einav et al. (2012). Leaving 401(k) investments aside for the moment, Einav et al. (2012) find (i) moderately positive correlations between contexts involving stakes of the same order of magnitude (small/small or moderate/moderate) and (ii) weakly positive correlations between contexts involving stakes of adjacent orders of magnitude (small/moderate). We corroborate the first find-

¹³We note that in an effort to establish the comparability of the choices they study, Einav et al. (2012, p. 2616) argue that "the incremental decisions across each domain are quite comparable in *expected* magnitude, . . . ranging from several hundred to a few thousand dollars" (emphasis added). We do not disagree. But the fact remains that the choices in their first three contexts differ categorically from the choices in their last three contexts in terms of the absolute magnitude of the stakes involved. Moreover, were we to classify choices according to the expected magnitudes of the options (or inter-option increments), this arguably would be inconsistent with taking a model-free approach, as it would presuppose a model that entails comparisons over expected values or utilities.

ing (for small/small) and extend it (to large/large) and progressively add a third: (iii) weakly negative correlations between contexts involving stakes of remote orders of magnitude (small/large). Returning to 401(k) investments, Einav et al. (2012) acknowledge that this context is "the most difficult to reconcile with any of the others" (p. 2636), and they attribute the difficulty to a difference in kind between investments and insurance. Our results suggest an alternative explanation: employees may perceive 401(k) investments as a borderline large-stakes context, particularly if they view their allocation choice as applying to more than just their current year's contributions. This could explain the extremely weak correlations (more or less zero) between 401(k) investments and every other context in Einav et al. (2012).

5 Discussion

We examine the hypothesis that risk preferences have a stable, context-invariant component using data on households' insurance choices. We study five insurance contexts, three involving small-stakes choices (deductibles) and two involving large-stakes choices (liability limits). Adopting the model-free approach of Einav et al. (2012), we assess the extent to which the households' choices display a stable ranking in their willingness to bear risk relative to their peers. While we find that the households' choices reflect a stable ranking in risk taking across the three small-stakes choices and across the two large-stakes choices, we also find that the households who take on more risk than their peers in small-stakes contexts tend to take on less risk than their peers in large-stakes contexts, and vice versa, which does not support the stability hypothesis.

What could explain our results? Four stories come readily to mind. None is unassailable, however, and so each leaves open questions for future research.

The first is a story about relative risk aversion. Suppose that rich households choose higher deductibles and higher liability limits than poor households. The intuition might be

that rich households want insurance against large losses but not small losses (which they can self-insure at a lower cost), whereas poor households want insurance against small losses but not large losses (because you can't get blood from a stone). This pattern of choices, which could explain our results, could arise from a population of households with standard EU preferences and heterogeneous relative risk aversion. Standard EU preferences feature a concave utility function that is defined over wealth and exhibits DARA (Pratt 1964; Arrow 1971).¹⁴ Concavity implies a positive willingness to pay for insurance. Let π denote this willingness. DARA implies that π decreases with wealth, which could account for rich households choosing higher deductibles than poor households. If the utility function also exhibits IRRA/CRRA/DRRA,¹⁵ then, ceteris paribus, π is increasing/constant/decreasing in stakes (Menezes and Hanson 1970; Zeckhauser and Keeler 1970). Thus, the right kind of heterogeneity in relative risk aversion (e.g., rich households have IRRA and poor households have CRRA) could account for rich households also choosing higher liability limits than poor households.

This story, while plausible, has at least two important counterpoints. The first is our analysis in Section 4.2.2, which casts doubt on the possibility that wealth differences are behind our results. The second is the Rabin (2000) critique, which contends that EU theory is not a plausible model of risk aversion across small- and large-stakes gambles.¹⁶

A second story features consumption commitments (i.e., spending obligations that are costly to adjust). Suppose that some households have consumption commitments while others do not. Chetty and Szeidl (2007) show that, within an EU framework, consumption commitments can induce non-concavities in the utility function (cf. Friedman and Savage 1948; Markowitz 1952) that increase risk aversion over small- and moderate-stakes gam-

¹⁴In this paragraph, DARA stands for decreasing absolute risk aversion, and IRRA, CRRA, and DRRA stand for increasing, constant, and decreasing relative risk aversion, respectively.

¹⁵For instance, Saha's (1993) expo-power utility function can exhibit DARA/IRRA or DARA/DRRA, while the power utility function exhibits DARA/CRRA.

¹⁶We note that the Rabin critique is not directly related to our main finding of rank instability of risk preferences across small- and large-stakes contexts.

bles relative to large-stakes gambles. Hence, the right kind of heterogeneity in consumption commitments (e.g., committed households have lower risk aversion over large-stakes gambles than other households) could generate a pattern of choices in which committed households choose lower deductibles and lower liability limits than other households, which could explain our results.

Again, this story, while plausible, is complicated by our sensitivity analysis. One implication of Chetty and Szeidl's (2007) theory is that consumption commitments "have a larger effect on risk aversion when agents are borrowing constrained" (p. 850). It follows that if heterogeneity in consumption commitments were driving our results, we would expect them to be sensitive to differences in access to credit. Our analysis in Section 4.2.3, however, suggests they are not.¹⁷

Probability distortions headline a third possible story. Suppose that households' subjective beliefs (in a subjective EU model) or decision weights (in a rank-dependent EU model) do not correspond to the objective risks. The right kind of heterogeneity in such beliefs or weights could explain our results. For example, suppose that some households overweight loss probabilities in large-stakes gambles but not small-stakes gambles, while other households overweight loss probabilities in small-stakes gambles but not large-stakes gambles. This could lead the former households to choose higher deductibles and higher liability limits than the latter households. Alternatively, suppose that some households grossly overweight loss probabilities in small-stakes gambles and mildly overweight them in large-stakes gambles, while other households do not overweight loss probabilities in any gambles (cf. Fehr-Duda et al. 2010).¹⁸ If in addition the former households are low risk while the other households are high risk, this could lead the former to choose lower deductibles and lower liability limits than the latter.

¹⁷Although Chetty and Szeidl (2007) adopt an EU framework, the non-concavities of the utility function insulate their model from the Rabin critique.

¹⁸Fehr-Duda et al. (2010) present evidence of this kind of pattern, but only for gambles in the gain domain. They find no substantial difference in stake-dependent probability weighting for gambles in the loss domain (like insurance).

The issue with each version of this story is that it requires a peculiar heterogeneity structure. (Indeed, we could level this criticism against the first two stories as well.) We are not aware of any empirical or theoretical support for the kind of heterogeneity—including, in some versions, the correlation between probability distortions and risk types—that is required by this story.

A final story involves limited consideration. Suppose that households do not consider every option when making their choices.¹⁹ In particular, suppose that some households consider only high deductibles and high liability limits, while other households consider only low deductibles and low liability limits. This pattern of heterogeneous limited consideration, which could explain our results, could arise from a variety of consideration set formation processes. For instance, it could arise from a sequential search process in which some households consider options in decreasing order (high to low), while other households consider options in increasing order (low to high), and all households generally stop searching before considering every option (cf. Weitzman 1979; Honka, Hortaçsu, and Wildenbeest 2019). Such a process could be optimal if, for example, option values and search costs are negatively correlated for some households and positively correlated for others. And such correlations could exist if, for instance, the company initially quotes only high liability limits to some households and low liability limits to others (based on the extent to which they are judgment proof), and then initially quotes only high deductibles to households that choose high liability limits (to decrease moral hazard) and low deductibles to households that choose low liability limits (to increase total premium).

In future research it would be worthwhile to further probe these and other potential explanations of our results. It would also be worthwhile to explore whether similar results obtain in other comparable datasets.

¹⁹Since at least Simon (1955), the economics literature has recognized that consumers (rationally) may not consider every available option when making decisions. Moreover, there is evidence in the literature of limited consideration in insurance contexts (e.g., Honka 2014; Barseghyan et al. 2021).

References

- 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): 137–160.
- Arrow, Kenneth J.** 1971. *Essays on the Theory of Risk Bearing*. Chicago: Markham Publishing.
- Baker, Scott R.** 2018. “Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data.” *Journal of Political Economy*, 126 (4): 1504–1557.
- Barksy, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro.** 1997. “Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study.” *Quarterly Journal of Economics*, 112 (2): 537–579.
- Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum.** 2021. “Heterogeneous Choice Sets and Preferences.” *Econometrica*, 89 (5): 2015–2048.
- Barseghyan, Levon, Francesca Molinari, and Joshua C. Teitelbaum.** 2016. “Inference under Stability of Risk Preferences.” *Quantitative Economics*, 7 (2): 367–409.
- Barseghyan, Levon, Jeffrey Prince, and Joshua C. Teitelbaum.** 2011. “Are Risk Preferences Stable Across Contexts? Evidence from Insurance Data.” *American Economic Review*, 101 (2): 591–631.
- Binswanger, Hans P.** 1980. “Attitudes toward Risk: Experimental Measurement in Rural India.” *American Journal of Agricultural Economics*, 62 (3): 395–407.

- Charness, Gary, Thomas Garcia, Theo Offerman, and Marie Claire Villeval.** 2020. “Do Measures of Risk Attitude in the Laboratory Predict Behavior under Risk in and outside of the Laboratory?” *Journal of Risk and Uncertainty*, 60 (2): 99–123.
- Chetty, Raj and Adam Szeidl.** 2007. “Consumption Commitments and Risk Preferences.” *Quarterly Journal of Economics*, 122 (2): 831–877.
- Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv.** 2007. “Consistency and Heterogeneity of Individual Behavior under Uncertainty.” *American Economic Review*, 97 (5): 1921–1938.
- Collier, Benjamin L., Daniel Schwartz, Howard C. Kunreuther, and Erwann O. Michel-Kerjan.** 2018. “Risk Preferences in Small and Large Stakes: Evidence from Insurance Contract Decisions.” National Bureau of Economic Research Working Paper 23579.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner.** 2011. “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences.” *Journal of the European Economic Association*, 9 (3): 522–550.
- Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen.** 2012. “How General are Risk Preferences? Choice under Uncertainty in Different Domains.” *American Economic Review*, 102 (6): 2606–2638.
- Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert.** 2010. “Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size.” *Journal of Risk and Uncertainty*, 40 (2): 147–180.
- Friedman, Milton and L. J. Savage.** 1948. “The Utility Analysis of Choices Involving Risk.” *Journal of Political Economy*, 56 (4): 279–304.

- Galizzi, Matteo M., Sara R. Machado, and Raffaele Miniaci.** 2016. “Temporal Stability, Cross-Validity, and External Validity of Risk Preferences Measures: Experimental Evidence from a UK Representative Sample.” <http://eprints.lse.ac.uk/67554/>.
- Galizzi, Matteo M., Marisa Miraldo, and Charitini Stavropoulou.** 2016. “In Sickness but Not in Wealth: Field Evidence on Patients’ Risk Preferences in Financial and Health Domains.” *Medical Decision Making*, 36 (4): 503–517.
- Handel, Ben, Igal Hendel, and Michael D. Whinston.** 2015. “Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk.” *Econometrica*, 83 (4): 1261–1313.
- Handel, Benjamin R.** 2013. “Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts.” *American Economic Review*, 103 (7): 2643–2682.
- Hausman, Daniel M.** 2012. *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.
- Holt, Charles A. and Susan K. Laury.** 2002. “Risk Aversion and Incentive Effects.” *American Economic Review*, 92 (5): 1644–1655.
- Honka, Elisabeth.** 2014. “Quantifying Search and Switching Costs in the US Auto Insurance Industry.” *RAND Journal of Economics*, 45 (4): 847–884.
- Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest.** 2019. “Empirical Search and Consideration Sets.” In *Handbook of the Economics of Marketing, Volume 1*, edited by Jean-Pierre Dubé and Peter Rossi, 193–257. Amsterdam: Elsevier.
- Ioannou, Christos A. and Jana Sadeh.** 2016. “Time Preferences and Risk Aversion: Tests on Domain Differences.” *Journal of Risk and Uncertainty*, 53 (1): 29–54.
- Jaffe, Sonia and Anup Malani.** 2018. “The Welfare Implications of Health Insurance.” National Bureau of Economic Research Working Paper 24851.

- Jaspersen, Johannes, Marc A. Ragin, and Justin R. Sydnor.** 2022. "Predicting Insurance Demand from Attitudes." *Journal of Risk and Insurance*, 89 (1): 63–96.
- Kachelmeier, Steven J. and Mohamed Shehata.** 1992. "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China." *American Economic Review*, 82 (5): 1120–1141.
- Markowitz, Harry.** 1952. "The Utility of Wealth." *Journal of Political Economy*, 60 (2): 151–158.
- Menezes, C. F. and D. L. Hanson.** 1970. "On the Theory of Risk Aversion." *International Economic Review*, 11 (3): 481–487.
- Morris, Darcy Steeg, Daniel Schwarcz, and Joshua C. Teitelbaum.** 2017. "Do Credit-Based Insurance Scores Proxy for Income in Predicting Auto Claim Risk?" *Journal of Empirical Legal Studies*, 14 (2): 397–423.
- Pratt, John W.** 1964. "Risk Aversion in the Small and in the Large." *Econometrica*, 32 (1/2): 122–136.
- Rabin, Matthew.** 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68 (5): 1281–1292.
- Saha, Atanu.** 1993. "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion." *American Journal of Agricultural Economics*, 75 (4): 905–913.
- Simon, Herbert A.** 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69 (1): 99–118.
- Tversky, Amos and Daniel Kahneman.** 1986. "Rational Choice and the Framing of Decisions." *Journal of Business*, 59 (4): S251–S278.

Weitzman, Martin L. 1979. "Optimal Search for the Best Alternative." *Econometrica*, 47 (3): 641–654.

Wolf, Charles and Larry Pohlman. 1983. "The Recovery of Risk Preferences from Actual Choices." *Econometrica*, 51 (3): 843–850.

Zeckhauser, Richard and Emmett Keeler. 1970. "Another Type of Risk Aversion." *Econometrica*, 38 (5): 661–665.

Table 1—Demographics and Claims

	Mean	Std dev	5th pctl	95th pctl
Auto policies				
Driver 1 age (years)	57	15	32	80
Driver 1 female	0.38	0.49		
Driver 1 single	0.20	0.40		
Driver 1 married	0.58	0.49		
Driver 2 indicator	0.43	0.50		
Driver 3+ indicator	0.03	0.16		
Vehicle 1 age (years)	5	3	1	11
Vehicle 2 indicator	0.48	0.50		
Vehicle 3+ indicator	0.03	0.17		
Insurance score	788	106	602	957
Collision claims (per annum)	0.089	0.286	0.000	0.600
Comprehensive claims (per annum)	0.024	0.125	0.000	0.000
Single limit claims (per annum)	0.085	0.277	0.000	0.597
Home policies				
Home age (years)	44	31	2	105
Home value (thousands of dollars)	213	155	90	430
Insurance score	733	100	562	888
All perils claims (per annum)	0.058	0.192	0.000	0.451
Personal liability claims (per annum)	0.002	0.027	0.000	0.000

Notes: The table reports descriptive statistics for the baseline sample of 2,690 households. Insurance scores in auto and home are based on information contained in credit reports.

Table 2—Choices and Prices

	Share (percentage)	Premium saving relative to safest option (dollars)			
		Mean	Std dev	5th pctl	95th pctl
Auto collision					
\$100	1.0				
\$200	15.2	40	23	15	84
\$250	11.6	80	46	31	168
\$500	63.8	134	77	52	281
\$1,000	8.3	174	100	67	365
Auto comprehensive					
\$50	5.1				
\$100	4.7	45	32	15	93
\$200	34.9	67	48	23	140
\$250	11.2	74	53	26	155
\$500	39.3	104	75	36	217
\$1,000	4.8	127	91	43	264
Home all perils					
\$100	0.3				
\$250	22.3	186	156	83	403
\$500	54.9	248	207	110	529
\$1,000	21.0	330	275	146	694
\$2,500	1.3	391	326	176	820
\$5,000	0.3	463	386	206	1001
Auto single limit					
\$60,000	0.2	109	46	55	200
\$100,000	8.6	102	43	52	189
\$200,000	0.7	78	33	40	143
\$300,000	43.9	68	29	34	125
\$500,000	43.0	57	24	29	106
\$1,000,000	3.6				
Home personal liability					
\$100,000	9.6	42			
\$200,000	0.8	32			
\$300,000	47.6	24			
\$400,000	0.2	19			
\$500,000	36.4	16			
\$1,000,000	5.4				

Notes: The table summarizes the coverage choices and pricing menus for the baseline sample of 2,690 households. Share is the percentage of households who chose a given option (deductible or limit, as the case may be). The safest option is the lowest deductible in the property coverages and the highest limit in the liability coverages.

Table 3—Baseline Results

	Auto collision	Auto comprehensive	Home all perils	Auto single limit
<i>Panel A. Spearman rank correlations</i>				
Auto comprehensive	0.617			
Home all perils	0.395	0.383		
Auto single limit	-0.129	-0.108	-0.224	
Home personal liability	-0.206	-0.219	-0.339	0.563
<i>Panel B. Correlation estimates from probit regression</i>				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.104	-0.056	-0.134	
Home personal liability	-0.149	-0.133	-0.205	0.574
<i>Panel C. Correlation estimates from linear regression</i>				
Auto comprehensive	0.552			
Home all perils	0.290	0.263		
Auto single limit	-0.077	-0.055	-0.113	
Home personal liability	-0.121	-0.114	-0.163	0.437

Notes: The table provides results for the baseline sample of 2,690 households. Each cell reports a pairwise correlation coefficient. For each correlation coefficient, the p-value associated with a test of whether the coefficient is different from zero is less than 0.01. The only exception is the correlation coefficient between auto comprehensive and auto single limit in panel B, for which the associated p-value is 0.023. The probit and linear regressions include controls for prices and risk.

Table 4—Sensitivity Analysis

	Auto collision	Auto comprehensive	Home all perils	Auto single limit
<i>Panel A. Correlation estimates accounting for umbrella choices</i>				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.134	-0.103	-0.135	
Home personal liability	-0.165	-0.150	-0.176	0.842
<i>Panel B. Correlation estimates with control for weath</i>				
Auto comprehensive	0.703			
Home all perils	0.399	0.336		
Auto single limit	-0.106	-0.060	-0.144	
Home personal liability	-0.151	-0.138	-0.214	0.570
<i>Panel C. Correlation estimates with controls for insurance scores</i>				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.105	-0.061	-0.136	
Home personal liability	-0.148	-0.132	-0.204	0.576
<i>Panel D. Correlation estimates with same-day choice window</i>				
Auto comprehensive	0.707			
Home all perils	0.446	0.375		
Auto single limit	-0.137	-0.103	-0.157	
Home personal liability	-0.133	-0.159	-0.203	0.649

Notes: Panels A, B, and C provide results for the baseline sample of 2,690 households. Panel D provides results for the subsample of 1,694 households who purchased all five coverages on the same day. Each cell reports a pairwise correlation coefficient estimated from a system of ordered probits with controls for prices and risk. In panel A, households who purchased umbrella coverage are treated as having chosen a new "highest limit" option in both auto single limit and home personal liability. In panel B, the probit regression includes an additional control for wealth. In panel C, the probit regression includes additional controls for insurance scores in auto and home. For each correlation coefficient, the p-value associated with a test of whether the coefficient is different from zero is less than 0.01. The only exception is the correlation coefficient between auto comprehensive and auto single limit in panels B and C, for which the associated p-value is 0.015.

Online Appendix:

Different Contexts, Different Risk Preferences?

Levon Barseghyan Joshua C. Teitelbaum Lin Xu

Draft: September 2, 2022

1 Risk Controls

The risk controls are expected annual claims under each coverage based on separate Poisson-gamma Bayesian credibility models. More specifically, we assume that household i 's claims under coverage j in year t follow a Poisson distribution with arrival rate λ_{ijt} . We treat λ_{ijt} as a latent random variable and assume that $\ln \lambda_{ijt} = \mathbf{z}'_{ijt} \boldsymbol{\alpha}_j + \epsilon_{ij}$, where \mathbf{z}_{ijt} is a vector of observables, $\boldsymbol{\alpha}_j$ is a vector of coefficients, ϵ_{ij} is an iid error term, and $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance ϕ_j . Utilizing our full dataset, we perform separate Poisson panel regressions with random effects to obtain maximum likelihood estimates of $\boldsymbol{\alpha}_j$ and ϕ_j for each coverage j . For each household i in the baseline sample, we then calculate the expected number of claims $\widehat{\lambda}_{ij}$ for each coverage j , conditional on the household's ex ante characteristics \mathbf{z}_{ij} and ex post claims experience $\boldsymbol{\gamma}_{ij}$, as follows: $\widehat{\lambda}_{ij} = \exp(\mathbf{z}'_{ij} \widehat{\boldsymbol{\alpha}}_j) E(\exp(\epsilon_{ij}) | \boldsymbol{\gamma}_{ij})$, where $E(\exp(\epsilon_{ij}) | \boldsymbol{\gamma}_{ij})$ is calculated assuming $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance $\widehat{\phi}_j$. By construction, $\widehat{\lambda}_{ij}$ takes into account both the systematic and idiosyncratic components of a households' risk type.¹

¹We refer to the above-described model as a Bayesian credibility model because $\widehat{\lambda}_{ij}$ corresponds to the Bayesian credibility premium in the actuarial literature (Denuit et al. 2007, ch. 3).

2 Classification of the Contexts in Einav et al. (2012)

In order to compare our results with those of Einav et al. (2012), we classify each of their contexts according to the magnitude of the stakes involved. Moreover, we apply the same criteria to classify their contexts that we use to classify our contexts. Recall that in our small-stakes contexts the values of the options and the inter-option increments range in the hundreds and thousands dollars, whereas in our large-stakes contexts the value of the options range in the hundreds of thousands and millions of dollars with inter-option increments that range in the hundreds of thousands dollars.

Based on their description of the coverage options in each context (Einav et al. 2012, pp. 2612-2616),² we conclude that none of their contexts involve large-stakes choices. Three of their six contexts—health, drug, and dental insurance—involve small-stakes choices. In health insurance, employees effectively choose among deductible options that range from zero to \$3,000 (with a mean inter-option increment of \$750) for in-network care and from \$500 to \$6,000 (with a mean inter-option increment of \$1,375) for out-of-network care. In drug insurance, employees choose among brand drug cost-sharing percentages that range from 30 percent to 50 percent for retail purchases and from 20 percent to 40 percent for mail-order purchases. The mean of the resulting annual drug claims is approximately \$1,500 and the 95th percentile is approximately \$5,500. In dental insurance, employees effectively choose between a maximum annual benefit of \$1,000 or \$2,000.

In two of the three remaining contexts—short-term disability insurance and 401(k) investments—the stakes range in the thousands and tens of thousands of dollars but not the hundreds of thousands of dollars, and so we classify them as moderate-stakes contexts. In short-term disability insurance, which replaces lost wages due to disability for up to six months, employees choose among wage-replacement rates that range from 60 percent to 100

²See also pp. 4-5 in their Online Appendix.

percent.³ The mean annual wage of the employees in their baseline sample is approximately \$58,000 and the 95th percentile is approximately \$114,000. At the mean claim duration, which Einav et al. (2012) report is approximately two months, this suggests that the value of the benefit ranges approximately from \$5,800 to \$9,700 for the average employee and does not exceed \$19,000 for 95 percent of employees. Even at the maximum claim duration, the value of the annual benefit ranges approximately from \$17,000 to \$29,000 for the average employee and does not exceed \$57,000 for 95 percent of employees. In 401(k) investments, contributing employees choose how to allocate their contributions among 13 different funds whose prospective monthly returns range from -11.69 percent to 16.79 percent.⁴ The mean annual contribution is approximately \$4,600 and the maximum allowable is \$18,000,⁵ with Alcoa matching contributions up to six percent. This suggests that the stakes range approximately from $-\$2,200$ to $\$8,300$ for the average contributor and from $-\$8,500$ to $\$32,400$ for all contributors.

We also classify the remaining context—long-term disability insurance—as involving moderate-stakes choices, though the classification is less straightforward than in the other contexts. Alcoa’s long-term disability plan replaces lost wages due to disability for durations longer than six months, subject to a six-month elimination period.⁶ Employees choose among three wage-replacement rates: 50 percent, 60 percent, or 70 percent. At the mean claim duration, which Einav et al. (2012) report is approximately one year,⁷ this suggests

³In their Appendix Table A1, Einav et al. (2012) note that "sometimes" the wage-replacement rates in short-term disability insurance range instead from 40 percent to 80 percent.

⁴Einav et al. (2012) abstract from the employees decisions as to whether and how much to contribute, but rather focus on how contributing employees choose to allocate their contributions across the funds. The range of monthly returns is taken from Appendix Table A2 in Einav et al. (2012), which reports summary statistics of the funds’ monthly returns from August 2005 to December 2007.

⁵Einav et al. (2012) state that the choices were made in 2004. We assume they reflect benefit elections for 2005. In 2005, the annual contribution limit was \$14,000 for employees under age 50 and \$18,000 for older employees.

⁶The elimination period is the period of time between the onset of disability and the time at which the employee is eligible to receive benefits.

⁷Einav et al. (2012) note that their claims data are truncated at about two years, which suggests the mean claim duration may be longer than one year. In a recent study of employer-provided long-term disability insurance, Autor, Duggan, and Gruber (2014) report a mean claim duration of 1.55 years and a

that the value of the benefit ranges approximately from \$29,000 to \$41,000 for the average employee and does not exceed \$80,000 for 95 percent of employees. At the maximum claim duration, which we assume could be as long as 45 years,⁸ the present value of the benefit could range into the hundred of thousands of dollars; but even in this extreme case the present value of the inter-option increments would range in the tens of thousands of dollars.⁹ All things considered, we conclude the stakes are best classified as moderate.

References

Autor, David, Mark Duggan, and Jonathan Gruber. 2014. "Moral Hazard and Claims Deterrence in Private Disability Insurance." *American Economic Journal: Applied Economics*, 6 (4): 110–141.

Denuit, Michel, Xavier Maréchal, Sandra Pitrebois, and Jean-François Walhin. 2007. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Chichester: John Wiley.

median of one year. Their sample consists of approximately 8 million quarterly observations from nearly 10,000 unique employers, and their claims data span eight years.

⁸Einav et al. (2012) do not report the maximum claim duration (or the 95th percentile) in their data, nor do they report the maximum benefit period under Alcoa's long-term disability plan. The maximum benefit period under many long-term disability plans is 2, 5, or 10 years, but under the most generous plans it runs until the employee's social security full retirement age, which is 67 for employee's born in 1960 or later. Assuming that Alcoa's plan has the most generous maximum benefit period and that its youngest eligible employee is 22 years old, we arrive at the assumption that the maximum claim duration could be as long as 45 years.

⁹We are assuming annual discount rates well in excess of 10 percent, which is consistent with the preponderance of the empirical evidence on time preferences (Frederick, Lowenstein, and O'Donohue 2002, pp. 377-380). For instance, Warner and Pleeter (2001) estimate the personal discount rates of approximately 66,000 U.S. military personnel who were offered separation benefits that consisted of a choice between a lump sum or an annuity, where the break-even discount rate was at least 17.5 percent. They find that "over half of the officers and over 90 percent of enlisted personnel chose the lump-sum payment, implying that the vast majority of personnel had discount rates of at least 18 percent" (p. 33). Based on regression analysis, they report mean discount rates of between 10 percent and 19 percent for officers and between 35 percent and 54 percent for enlisted personnel, depending on the model specification (p. 48, tbl. 6). As Frederick, Lowenstein, and O'Donohue (2002, p. 385) note, this field study "is particularly compelling in terms of credibility of reward delivery, magnitude of stakes, and number of subjects."

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen. 2012. “How General are Risk Preferences? Choice under Uncertainty in Different Domains.” *American Economic Review*, 102 (6): 2606–2638.

Frederick, Shane, George Lowenstein, and Ted O’Donohue. 2002. “Time Discounting and Time Preference: A Critical Review.” *Journal of Economic Literature*, 40 (2): 351–401.

Warner, John T. and Saul Pleeter. 2001. “The Personal Discount Rate: Evidence from Military Downsizing Programs.” *American Economic Review*, 91 (1): 33–53.