

Department of Economics & Finance

Risk Attitudes, Sample Selection and Attrition in a Longitudinal Field Experiment

Glenn W. Harrison, Morten I. Lau and Hong Il Yoo

Working Paper No. 07, 2017

Department Economics and Finance
Durham University Business School
Mill Hill Lane
Durham DH1 3LB, UK
Tel: +44 (0)191 3345200
<https://www.dur.ac.uk/business/research/economics/>

© Durham University, 2017

Risk Attitudes, Sample Selection and Attrition in a Longitudinal Field Experiment

by

Glenn W. Harrison, Morten I. Lau and Hong Il Yoo [†]

July 2017

Abstract. Longitudinal experiments allow one to evaluate the temporal stability of latent preferences, but raise concerns about sample selection and attrition that may confound inferences about temporal stability. We evaluate the hypothesis of temporal stability in risk preferences using a remarkable data set that combines socio-demographic information from the Danish Civil Registry with information on risk attitudes from a longitudinal field experiment. Our experimental design builds in explicit randomization on the incentives for participation. The results show that the use of different participation incentives can affect sample response rates and help one identify the effects of selection. Correcting for endogenous sample selection and panel attrition changes inferences about risk preferences in an economically and statistically significant manner. We draw mixed conclusions on temporal stability of risk preferences that depend on which aspect of temporal stability one is interested in.

[†] Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); Copenhagen Business School, Denmark (Lau), and Durham University Business School, UK (Lau and Yoo). E-mail contacts: gharrison@gsu.edu, mla.eco@cbs.dk and h.i.yoo@durham.ac.uk. Harrison is also affiliated with the School of Commerce, University of Cape Town. We thank the Danish Social Science Research Council for research support under project 275-08-0289 and 12-130950, the Carlsberg Foundation under grant 2008-01-0410, and John DiNardo, Justin McCrary and Bill Schworm for helpful comments.

Table of Contents

1. Introduction.	-1-
2. Data.	-8-
A. Field Sampling Procedures.	-8-
B. Experiments to Infer Risk Attitudes.	-11-
3. Identification of Risk Preferences.	-12-
A. Baseline EUT Specification.	-13-
B. EUT Specification with Endogenous Sample Selection and Panel Attrition.	-16-
C. Rank Dependent Utility Theory Specifications.	-21-
4. Results.	-22-
A. Temporal Stability of Risk Attitudes.	-24-
B. Effects of Sample Selection and Attrition on Risk Attitudes under EUT.	-26-
C. Effects of Sample Selection and Attrition on Risk Attitudes under RDU.	-27-
D. Incorporating Observed Heterogeneity.	-30-
5. Conclusions.	-31-
References.	-44-
Appendix A: Instructions (WORKING PAPER).....	-49-
Appendix B: Additional Estimations (WORKING PAPER).....	-51-

1. Introduction

Any longitudinal survey or experimental design raises concerns about sample selection and attrition, and response rates may vary dramatically depending on the nature of the study and incentives provided in the design. Controlling for endogenous effects of sample selection requires some background information on subjects who did not select into the survey or experiment, so that one can estimate a latent selection process and its correlation with a primary outcome of interest. This information is often missing, and most longitudinal studies are concerned just with attrition effects. For non-participants, attrition outcomes are also missing, and strictly speaking one cannot control for attrition effects without addressing endogenous selection first. Without controlling for selection effects, the estimates of a latent attrition process may be subject to selection bias even when there is no effect of selection on the primary outcome.

Using a structural model of risky choices which allows for endogenous sample selection *and* panel attrition, we analyze data from a longitudinal field experiment with a stratified sample of the adult Danish population. The data are linked to administrative data from the Civil Registry in Denmark, allowing us to observe background information on non-participants. We illustrate the importance of controlling for within-wave and between-wave effects of sample selection in the evaluation of individual risk attitudes at different points in time.

Temporal stability of risk preferences is a common assumption in evaluations of economic behavior.¹ Testing this assumption with the same individuals requires, of course, that one address problems of sample selection and attrition. We design and evaluate a longitudinal field experiment with a nationally representative sample of Danish adults between 19 and 75 years of age to address this question, and provide a range of findings on temporal stability of risk preferences. The sample is randomly drawn from the Civil Registry and stratified with respect to population size in each county.

¹ The term stability can mean unconditional stability or it can mean stable preferences conditional on a given set of covariates. In the latter case the question is whether preferences are a stable (and known) *function* of those covariates (Andersen, Harrison, Lau and Rutström [2008; §2]). We consider both forms of stability.

Our design builds in explicit randomization on the incentives for participation, an idea suggested by the theoretical literature on sample selection models and easy to implement in the sampling process and subsequent experiment.

The classic problem of sample selection refers to possible recruitment biases, such that individuals with certain types of characteristics are more likely to be in the observed sample. The statistical problem is that there may be some unobserved characteristics which simultaneously affects someone's chance of being in the sample as well as affecting other outcomes that the analyst is interested in. In any longitudinal study, there is also an inherent scope for post-recruitment selection bias due to panel attrition, which occurs as some subjects may leave the panel.² We build on the direct likelihood approach of Heckman [1976], Hausman and Wise [1979] and Diggle and Kenward [1994] and use maximum simulated likelihood to estimate unique probit-kernel models that consider the full longitudinal design of the experiment. Our models control for the effects of selection and attrition on risk preferences inferred from both waves of the experiment, as well as addressing unobserved heterogeneity in risk preferences of the underlying population.

We consider a structural analysis of two theories of decision making under risk, specifically Expected Utility Theory (EUT) and Rank Dependent Utility (RDU), where the latter is a highly influential alternative to EUT. Each theory has a set of structural parameters that characterize risk preferences. Previous analyses of temporal stability do not control for recruitment bias, and focus either on population averages of the structural parameters or on individual-level estimates which have no structural link to the population distribution of risk preferences. In contrast, our analysis controls for endogenous sample selection and attrition, and captures unobserved heterogeneity around the population averages by modeling all structural parameters as individual-level random coefficients that follow a population distribution. We allow the population distribution to vary over

² The attrition problem is not the same as the dropout problem. As stressed by Heckman, Smith and Taber [1998], the latter refers to subjects that leave some randomized program or intervention, but that remain in the sample. The attrition problem concerns subjects that completely drop out of the sample.

time, and the random coefficients to be correlated with the error terms in the selection and attrition equations.

This estimation approach allows us to consider temporal stability of risk attitudes at two different levels, with and without controls for endogenous sample selection and attrition: (i) the population level, by comparing the population distributions of structural parameters over time, and (ii) the individual level, by considering the correlation between individual-specific random coefficients over time. Our direct likelihood approach is inspired by the trivariate probit model of Capellari and Jenkins [2004], which includes two different types of selection equations, but their primary outcome equation is the linear index probit model and their selection equations do not address selection bias in the sense of recruitment bias.³ We are not aware of past statistical models that capture unobserved heterogeneity in latent structural parameters with controls for recruitment bias and/or attrition bias in longitudinal studies.

No existing studies test temporal stability of risk attitudes in the context of a model that addresses unobserved preference heterogeneity across the population. Glöckner and Pachur [2012] and Zeisberger, Vrecko and Langer [2012] are so far the only studies that test temporal stability of risk preferences at the individual level. But they do not consider temporal stability at the population level and do not control for sample selection or attrition bias.⁴

³ Capellari and Jenkins [2004] analyze the transition of poverty states in the UK using a first-order Markov model. The primary outcome equation describes the present poverty state, and features parameters that depend on the initial poverty state. The two types of selection equations correct for endogenous selection into the initial poverty state and endogenous panel attrition.

⁴ Glöckner and Pachur [2012] and Zeisberger, Vrecko and Langer [2012] estimate one set of structural parameters for Cumulative Prospect Theory for each individual subject, and compare the point estimates over one-week and one-month time periods, respectively. Their statistical tests of temporal stability, however, do not fully account for random sampling variations in the estimates. Hey and Orme [1994] were the first to consider individual level estimation of latent risk attitudes, which requires a sufficiently large number of observations per subject; they had a sample of 80 subjects with 100 observations per subject. Later applications of individual level estimation of latent preferences also consider individual discount rates (Andersen, Harrison, Lau and Rutström [2014]) and intertemporal correlation aversion (Andersen, Harrison, Lau and Rutström [2017]). Harrison and Swarthout [2016] estimate the full set of latent structural parameters for Cumulative Prospect Theory at the individual level and find considerable variation in risk preferences. To control for endogenous sample selection and/or attrition bias and study temporal stability at the population level one must pool observations over all subjects and estimate the population distributions of individual level coefficients, which we do.

Existing studies on temporal stability of risk attitudes do not control for selection bias, and we are aware of only one study (Andersen, Harrison, Lau and Rutström [2008]) that controls for attrition bias. In fact, most studies do not even make a passing reference to “sample selection” and, perhaps more remarkably, “attrition” or “retention” (Smidt [1997], Goldstein, Johnson and Sharpe [2008], Baucellis and Villasís [2010], Glöckner and Pachur [2012] and Zeisberger, Vrecko and Langer [2012]). A recent study by Dasgupta, Gangadharan, Maitra and Mani [2017] reports a significant difference in the sample average risk attitudes of the attrited and the retained, but does not undertake statistical correction for attrition bias and does not mention selection bias.

We draw several conclusions from our statistical analysis. First, we find evidence that *the use of different fixed recruitment fees can affect the decision to participate in our experiment*.⁵ When we used a relatively substantial recruitment fee of 500 kroner, which is about 100 US dollars, 24.1% of invitees accepted the invitation to the initial wave of our experiment. The initial acceptance rate fell to 18.1% when we instead used 300 kroner. Of course, this is just a “law of demand” effect from paying more money for people to participate, but demonstrates that there are indeed deliberate decisions being made about participation. The second wave of our experiment paid the same recruitment fee of 300 kroner to every person, and there was no significant difference in the retention rates of subjects who were initially recruited with the high fee (48.4%) and subjects who were initially recruited with the low fee (54.7%).

Second, we find evidence that *correcting for endogenous sample selection and panel attrition changes our inferences about risk preferences in an economically and statistically significant manner*. The results suggest that one should not discount the potential effects of selection and attrition *a priori*, even when a self-selected sample and an underlying population of interest look more or less similar in terms of observed characteristics. Subjects participating in each wave of our experiments have demographic

⁵ Paying no fixed recruitment fee is not a panacea for the sample selection issues we consider: it just masks it, and makes it impossible to evaluate since there is no variation in those fees. There are other sensible reasons why one should avoid zero show-up fees, since that could generate altogether different, and nasty, biases in sample selection documented by Kagel, Battalio and Walker [1979] and Eckel and Grossman [2000].

characteristics that are comparable to the adult population in Denmark, but without correcting for endogenous selection and attrition our EUT specification would have overestimated the average Dane's relative risk aversion in the first wave by a factor of about 2.⁶ Under RDU, non-linear probability weighting, that captures pessimism or optimism in relation to objective probabilities, may generate a positive or negative risk premium even when the individual is risk-neutral in terms of the utility function. Without correction for endogenous selection and attrition, our RDU specification would have substantially underestimated the population share of individuals who have an "inverse-S" probability weighting function that captures optimism for small probabilities and pessimism for large probabilities.

Finally, we draw *mixed conclusions on temporal stability of risk preferences that depend on which aspect of temporal stability one is interested in*. The range of results reflect the strengths of our empirical specifications that allow us to define and test temporal stability in several ways. For example, consider risk aversion in the EUT sense of a concave utility function. Under both EUT and RDU, we find that the average Dane is risk averse in this sense, and this conclusion is robust over time. But we still find some instability in the population distribution of risk aversion, under EUT because the average Dane becomes more risk-averse over time, and under RDU because there is a decline in the extent of unobserved preference heterogeneity around the average. When focusing on the within-individual autocorrelation of risk aversion, we find estimates between 0.40 and 0.45, which is between the two extreme cases of completely unrelated and completely stable preferences. Of course, under RDU risk preferences are also characterized by the probability weighting function. We find more evidence on the stability of the probability weighting function than for the utility function, both at the population and individual levels.

⁶ Andersen, Harrison, Lau and Rutström [2008] analyzed the stability of risk preferences in the same population between June 2003 and November 2004. They find evidence of stable risk preferences. Harrison, Lau and Rutström [2005] focussed on the analysis of the first experiment in June 2003, and found that the average Dane was risk averse. However, both studies did not randomize incentives for participation and did not undertake correction for endogenous selection into the initial experiment. Nor did they consider unobserved preference heterogeneity and the possibility of probability weighting under RDU.

Our use of exogenously varied recruitment fees demonstrates how one can constructively design features of a survey or experiment to facilitate empirical identification of sample selection effects. Building on Heckman [1976][1979], the emphasis in the literature has been on the discovery of some “exclusion restrictions,” referring to variables that affect the probability of selection but do not affect the primary outcome of interest.⁷ The collection of these variables could be designed by the surveyor or experimenter, but often were not.⁸ In most cases analysts simply have to live with the existing set of variables in a survey or experiment, and search for exclusion restrictions on an *a priori* basis. The later theoretical literature, typified by Das, Newey and Vella [2003], stresses the value of direct controls over the probability of selection, rather than relying on some variables selected on an *a priori* basis.

We know of only two applications of this constructive approach to building exclusion restrictions into the experimental design.⁹ Each example made an important methodological step by operationalizing a controlled basis for inferring selection bias or attrition bias. Nevertheless, neither example had access to information on non-participants that we have, nor considered the interaction between sample selection and panel attrition as we do.

The first example is the Survey Supply Experiment, undertaken as a module of the Index of Hospital Quality survey. Philipson [2001] analyzed data from this experiment, in which 23% of

⁷ Without such “exclusion restrictions,” identification of sample selection models has to rely on the validity of functional form assumptions alone, such as the bivariate normality of the error terms in the maximum likelihood estimation of the standard Heckman model. Identification in this instance is formally achieved, but is known to be “weak” (Meng and Schmidt [1985] and Keane [1992]). Exclusion restrictions are formally required for identification when semi-parametric specifications are used (Lee [1995]).

⁸ It is folklore in survey research that information is retained on how many calls were made to a subject, how hard they were to contact in other ways, or which interviewer conducted the survey. Although not the object of randomization, information of this kind might be used as an instrument to model the probability of selection.

⁹ One may find more examples when focussing on conceptual plans instead of actual applications. For instance, in evaluating the serious effects of attrition on psychotherapy, Leon et al. [2006; p. 1004] noted in passing that a “... very simple, yet overlooked, strategy for dealing with the inevitable problem of dropout is to collect data that can help predict attrition.” What they had in mind, following Demirtas and Schafer [2003], was to ask subjects how likely it was that they would show up again, but they also raised the possibility of offsetting transportation or logistical costs (p. 1004), which is related to our design with differential financial incentives for participation.

potential participants were randomized to the treatment group that would receive 50 US dollars for returning the survey questionnaire, whereas the control group faced no such incentive. The financial incentive resulted in a higher response rate of 59.3% for the treatment group of 121 randomly selected physicians, in comparison with a response rate of 50% for the control group of 298 physicians. The estimated mean income of the physicians in the sample became 50% larger after correcting for selection bias. The missing information on non-participants, however, meant that the effects of selection were identified by some strong *ad hoc* assumptions about the effects of the financial incentive and survey response rates on the uncorrected mean income.¹⁰

The second example is the follow-up for the Longitudinal Movement to Opportunity (MTO) field experiment, in which 30% of the sample was randomly assigned to more intensive follow-up: see Orr et al. [2003; Exhibit B, §B1.3] and DiNardo, McCrary and Sanbonmatsu [2006]. This randomized follow-up was in addition to the primary randomization to treatment: (i) a housing voucher with some strings attached and some counseling, (ii) a housing voucher with no strings attached and no counseling, and (iii) a control group. This additional randomization to more intensive follow-up had virtually no effect on results, however, since the effective response rates for the long-term MTO follow-up were around 90% and similar across primary treatments (Sanbonmatsu et al. [2011; p. 259]).¹¹

A key feature of the inferential problem considered in our experiment is that the “outcome variable” of interest is a *latent* characteristic: risk aversion. The context is fundamentally different from the cases that Philipson [1997][2001] considered, initially in a thought experiment (Philipson

¹⁰ Specifically, it was assumed that the uncorrected mean income was an increasing function of the financial incentive (Philipson [2001, p. 1101]) and was linear in survey response rates (Philipson [2001, p. 1109]).

¹¹ In many respects a similar methodological approach is employed by Behaghel, Crépon, Gurgand and Le Barbanchon [2009]. They evaluate two independent surveys of virtually the same population of job seekers in France: one survey involved a long telephone survey and had a 50% response rate, and the other survey involved a short telephone survey, augmented by administrative data, and had a higher 80% response rate. Using non-parametric methods from Horowitz and Manski [2000] and Lee [2009], they show that the two surveys lead to dramatically different estimates of the effects of career counseling programs on job search outcomes, arguing that the first survey suffers from severe selection bias.

[1997, §3]) and later in an empirical analysis (Philipson [2001]), where one could use randomized recruitment fees to remove selection bias from the estimated mean of an *observable* characteristic. This also means that we cannot replace data from subjects exhibiting non-response with administrative data, as many studies have done to assess the seriousness of sample selection and attrition (e.g., Grasdahl [2001], Behaghel et al. [2009] and Ludwig et al. [2013]).

There is some evidence from clinical drug trials that persuading patients to participate in randomized studies is much harder than persuading them to participate in non-randomized studies (e.g., Kramer and Shapiro [1984; p.2742ff.]). The same problem applies to social experiments, as evidenced by the difficulties that can be encountered when recruiting decentralized bureaucracies to administer random treatments (e.g., Hotz [1992]). For example, Heckman and Robb [1985] note that the refusal rate in one randomized job training program was over 90%. With the renewed popularity of randomized control trials in social sciences, evaluation of the potential effects of “randomization bias” is urgent.¹² Our methods of controlling for endogenous sample selection and attrition have broader applications to randomized control trials that consider causal effects of treatments on latent variables of interest in economic policy, such as welfare effects (Harrison [2011]).

2. Data

A. Field Sampling Procedures

Between September 28 and October 22, 2009, we conducted an artefactual field experiment with 413 Danes.¹³ The sample was drawn to be representative of the adult population as of January 1, 2009, using sampling procedures that are virtually identical to those documented at length in Harrison, Lau, Rutström and Sullivan [2005]. We received a random sample of the population aged

¹² This is also true, of course, for the effects of attrition in general. Hausman and Wise [1979; p. 455ff.] note that attrition “...may negate the randomization in the initial experimental design. If the probability of attrition is correlated with experimental response, then traditional statistical techniques will lead to biased and inconsistent estimates of the experimental effect.”

¹³ An artefactual field experiment is defined by Harrison and List [2004] as involving the use of artefactual instructions, task and environment with a field subject pool.

between 18 and 75, inclusive, from the Danish Civil Registration Office, stratified the sample by geographic area, and sent out 1,996 invitations. We drew this sample of 1,996 invitees from a random sample of 50,000 adult Danes obtained from the Danish Civil Registration Office, which includes information on sex, age, residential location, marital status, and whether the individual is an immigrant. Thus we are in the fortunate, and rare, position of knowing some basic demographic characteristics of the individuals that do *not* agree to participate in our experiment.¹⁴

At a broad level our final sample is representative of the population: the sample of 50,000 subjects had an average age of 49.8, 50.1% of them were married, and 50.7% were female; our final sample of 413 subjects had an average age of 48.7, 56.5% of them were married, and 48.2% were female. We stress this comparison because it is often made to assuage concerns about sample selection: check if the final sample is similar to the population in a few observed characteristics, and then assume it is representative in all characteristics, including those that are latent and unobserved. In the absence of the type of data we have access to in Denmark, this is a reasonable “second best” procedure, but our results show that it may be an inadequate check on endogenous sample selection effects.

The initial recruitment letter for the experiment explained the purpose and that it was being conducted by Copenhagen Business School. The letter clearly identified that there would be fixed and stochastic earnings from participating in the survey. In translation, the uncertainty was explained as follows:

You can win a significant amount

To cover travel costs, you will receive 500 kroner at the end of the meeting. Moreover, each participant will have a 10 percent chance of receiving an amount between 50 and 4,500 kroner in one part of the survey. In another part of the survey, each participant will have a 10 percent chance of receiving at least 1,500 kroner. Some of these amounts will also be paid out at the end of the meeting, and some amounts will be paid out in the future. A random choice will decide who wins the money in the different parts of the survey.

¹⁴ It is possible to extend this list of characteristics by taking our experimental data to Statistics Denmark, which stores the same data that we obtained from the Civil Registration Office, and merging it with the entire set of data that is available on all of the invited subjects. One can then undertake the same statistical analyses but with a larger set of co-variates to explain sample selection.

The fixed amount is 500 kroner in the treatment that this text comes from, and 300 kroner in another treatment. Subjects were randomly assigned to one of these two recruitment treatments. The stochastic earnings referred to in the recruitment letter were for a risk aversion task and a separate task eliciting individual discount rates. Thus the subjects should have anticipated the use of randomization in the experiment.

The experiments were conducted in hotel meeting rooms around Denmark, so that travel logistics for the invited sample would be minimized. Various times of day were also offered to subjects, to facilitate a broad mix of attendance. The largest session had 15 subjects, but most had fewer. The procedures were standard: Appendix A (available online) documents an English translation of the instructions, and shows a typical screen display for the risk aversion task. Subjects were given written instructions which were read out and then made choices in a trainer task for small non-monetary rewards. The trainer task was “played out” and illustrated the procedures in the experiment. All decisions were made on computers. After all choices had been made the subject was asked a series of standard socio-demographic questions.

There were 40 risk attitude choices and 40 discounting choices, and each subject had a 10% chance of being paid for one choice in each block of 40 choices. The risk attitude choices preceded the discounting choices in one treatment, and *vice versa* in another treatment. Average payments for the risk attitude choices were 242 kroner, and the average payments for the discounting choices were 201 kroner (although some were for deferred receipt), for a combined average of 443 kroner. The exchange rate at the time was close to 5 kroner per U.S. dollar, so expected earnings from these tasks combined were \$91. The subjects were also paid a 300 kroner or 500 kroner fixed show-up fee, plus earnings from subsequent tasks.¹⁵

¹⁵ An extra show-up fee of 200 kroner was paid to 35 subjects who had received invitations stating 300 kroner, but then received a final reminder that accidentally stated 500 kroner. The additional tasks earned subjects an average of 659 kroner, so total earnings from choices made in the session averaged 1102 kroner, or roughly \$221, in addition to the fixed fee of \$60 or \$100.

Between April 2010 and October 2010 we repeated the risk aversion and discounting tasks with 182 of the 413 subjects who participated in the first experiment.¹⁶ Each subject was interviewed in private in the new experiment, and the meeting was conducted at a convenient location for them (e.g., their private residence or the hotel where the first experiment took place). All subjects were paid a fixed fee of 300 kroner for their participation in the second experiment.

Table 1 provides the sample response in each panel wave, and definitions of the explanatory variables used in the statistical analysis and summary statistics. We observe a significant difference in sample response with the high recruitment fee compared to the low recruitment fee. The drop from 24.1% to 18.1% in the first wave is statistically significant according to a Fisher Exact test, with a p -value less than 0.001. After participating in the first wave, the sample response to recruitment into the second wave was slightly lower for those recruited into the first wave with the high recruitment fee compared to those recruited with the low fee. The sample response rates were 48.4% and 54.7% in the second wave, and are not statistically different according to a Fisher Exact test with a two-sided p -value of 0.24. One might infer from these statistics that the effects of attrition on elicited risk attitudes are not significant, but of course that depends on who responded, which can only be assessed with an appropriate statistical model.

B. Experiments to Infer Risk Attitudes

Risk attitudes were evaluated from data in which subjects made a series of binary lottery choices. For example, lottery A might give the individual a 50-50 chance of receiving 1600 kroner or 2000 kroner to be paid today, and lottery B might have a 50-50 chance of receiving 3850 kroner or

¹⁶ There were four steps in the construction of this sub-sample. First, we divided the country into five regions, and each region was divided into sub-regions. Each sub-region was assigned 1 or 2 numbers, in rough proportionality to the population of the sub-region. In total we assigned 24 numbers. Second, although Denmark is a relatively small country, it was necessary to consider logistical constraints, and we randomly picked 12 of the 24 numbers for the experiment in April 2010 and the remaining 12 numbers for the experiment in October 2010. Third, we picked the first 50% of the randomly sorted records within each sub-region. This provided a sub-sample of 100 subjects for each experiment. Fourth, we contacted subjects by phone and invited them to participate again in the experiments.

100 kroner today. The subject picks A or B. We used the procedures of Hey and Orme [1994], and presented each binary choice to the subject as a “pie chart” showing prizes and probabilities. We gave subjects 40 choices, in four sets of 10 with the same prizes. The prize sets employed are: [A1: 2000 and 1600; B1: 3850 and 100], [A2: 1125 and 750; B2: 2000 and 250], [A3: 1000 and 875; B3: 2000 and 75] and [A4: 2250 and 1000; B4: 4500 and 50]. The order of these four sets was randomized for each subject, with the probabilities varying within each set.¹⁷ We refer to the first and last of these four prize sets as the “high stakes” lotteries compared to the second and third. Each subject saw both the high stakes and low stakes lottery sets. All subjects in the experiment were presented with the same set of decision tasks.

We asked each subject to respond to all 40 risk aversion tasks and then randomly decided which one to play out using numbered dice. The large incentives and budget constraints precluded us from paying all subjects, so each subject was given a 10% chance to actually receive the payment associated with his decision. The typical findings from lottery choice experiments of this kind are that individuals are generally averse to risk, and that there is considerable heterogeneity in risk attitudes across subjects: see Harrison and Rutström [2008] for an extensive review.

3. Identification of Risk Preferences

We first write out a structural model to estimate risk attitudes assuming EUT, to focus on essentials. We then discuss how the likelihood function changes to account for sample selection and attrition, and then finally discuss the extension from EUT to the more general RDU model.

¹⁷ Within each prize set the 10 choices were presented one at a time in an ordered manner, with the probability of the high prize starting at 0.1 and increasing by 0.1 until the last choice is between two certain amounts of money.

A. Baseline EUT Specification

Consider the estimation of risk preferences in the simplest possible model of decision-making under risk, EUT, without worrying about sample selection or attrition. In our experiment, each decision task presented a choice between two lotteries, and each lottery had two potential outcomes. Let M_{ij} be the j^{th} outcome of lottery i , where $i=A,B$ and $j=1,2$. Assume that the utility of an outcome is given by the constant relative risk aversion (CRRA) specification

$$U(M_{ij}) = M_{ij}^{(1-r)} / (1-r) \quad (1)$$

for $r \neq 1$, where r is the CRRA coefficient. Then, under EUT, $r=0$ denotes risk neutral behavior, $r>0$ denotes risk aversion, and $r<0$ denotes risk loving behavior.

EUT predicts that the observed choice is lottery B when it gives the larger expected utility (EU) than lottery A and *vice versa*. Probabilities for each outcome, $p(M_{ij})$, are those that are induced by the experimenter, so the EU of lottery i is simply the probability weighted average of its outcome utilities,

$$EU_i = p(M_{i1}) \times U(M_{i1}) + p(M_{i2}) \times U(M_{i2}), \quad (2)$$

where $p(M_{i2}) = 1 - p(M_{i1})$. Let y denote a binary indicator of whether the observed choice is lottery B ($y = 1$) or lottery A ($y = 0$). Using the indicator function $\mathbf{I}(\cdot)$, the observed choice under EUT can be compactly written as $y = \mathbf{I}[(EU_B - EU_A) > 0]$.

To allow observed choices to deviate from deterministic theoretical predictions, the EUT model is combined with a stochastic behavioral error term. Specifically, assume that the choice depends not only on the EU difference, but also on a random error term ϵ such that $y = \mathbf{I}[(EU_B - EU_A) + \epsilon > 0]$. Assume further that ϵ is normally distributed with the standard deviation of μ , $\epsilon \sim N(0, \mu^2)$. The choice probability of lottery B is then $\Phi(\nabla EU)$ where $\Phi(\cdot)$ is the standard normal cumulative density function (CDF), and the index ∇EU is given by

$$\nabla EU = (EU_B - EU_A) / \mu. \quad (3)$$

It follows that the likelihood function for each choice observation takes the form

$$P(r, \mu) = \Phi(\nabla EU)^y \times (1 - \Phi(\nabla EU))^{(1-y)}. \quad (4)$$

As the noise parameter μ approaches 0, this stochastic EUT specification collapses to the deterministic EUT model; conversely, as μ gets arbitrarily large, it converges to an uninformative model which predicts a 50:50 chance regardless of the underlying EU difference. This is one of several types of behavioral error stories that could be used (Wilcox [2008]).

To clarify our econometric methods, more notation is needed than one would typically see in the context of non-linear models for panel data. We subscript the choice-level likelihood function in equation (4) as $P_{ntw}(r_{nw}, \mu)$ henceforth, to emphasize that it describes subject n 's choice in decision task t of panel wave w .¹⁸ The CRRA coefficient r_{nw} is indexed by subject n and wave w for two reasons. First, to capture unobserved preference heterogeneity across individuals, we model the CRRA coefficient as an individual-specific random coefficient drawn from a population distribution of risk preferences. Second, to test temporal stability, we allow the underlying population distribution, as well as the CRRA coefficient drawn from it, to vary freely across waves. We use $f(r_{n1}, r_{n2}; \theta)$ to denote the joint density function for the random CRRA coefficients, where θ is a set of parameters that characterize their joint distribution.

It is possible to estimate the set of parameters θ directly and draw inferences about the population distribution of risk preferences, once the joint density $f(r_{n1}, r_{n2}; \theta)$ is fully specified. Assume that r_{n1} and r_{n2} are jointly normal so that $\theta = (\bar{r}_1, \bar{r}_2, \sigma_{r1}, \sigma_{r2}, \sigma_{r1r2})$, where $\bar{r}_w$ and σ_{rw} are the population mean and standard deviation of the CRRA coefficient r_{nw} , and σ_{r1r2} is the covariance between r_{n1} and r_{n2} . Conditional on a particular pair of CRRA coefficient draws, the likelihood of observing a series of 40 or 80 choices made by subject n can be specified as

$$\begin{aligned} CL_n(r_{n1}, r_{n2}, \mu) &= \prod_t P_{nt1}(r_{n1}, \mu) && \text{if } s_{n2} = 0 \\ &= \prod_t P_{nt1}(r_{n1}, \mu) \times \prod_t P_{nt2}(r_{n2}, \mu) && \text{if } s_{n2} = 1 \end{aligned} \quad (5)$$

¹⁸ We repeated the same set of experiments across two panel waves, and within each wave the subject completed a series of decision tasks over 40 lottery pairs. The outcomes and probabilities associated with lottery pairs vary from task to task, and the same subject may make different choices across tasks and waves. Each lottery outcome and its probability are then M_{ijntw} and $p(M_{ijntw})$, leading to the expected utilities EU_{intw} and the index function ∇EU_{ntw} . The indicator y_{ntw} is 1 (0) if subject n chooses lottery B (lottery A) in decision task t of the experiment in wave w .

where s_{n2} is an indicator of whether subject n participated in only the first panel wave ($s_{n2} = 0$) or both panel waves ($s_{n2} = 1$). Since r_{n1} and r_{n2} are modeled as random coefficients, the “unconditional” (Train [2009, p.146]) or actual likelihood of subject n ’s choices is then obtained by taking the expected value of $CL_n(r_{n1}, r_{n2}, \mu)$ over the joint density $f(r_{n1}, r_{n2}; \theta)$

$$L_n(\bar{r}_1, \bar{r}_2, \sigma_{r1}, \sigma_{r2}, \sigma_{r1r2}, \mu) = L_n(\theta, \mu) = \iint CL_n(r_{n1}, r_{n2}, \mu) f(r_{n1}, r_{n2}; \theta) dr_{n1} dr_{n2}. \quad (6)$$

Unobserved heterogeneity is similarly integrated out from many textbook models for panel data, such as random effects probit (Wooldridge [2010, p.613]). Our application is distinctive because unobserved heterogeneity enters the index function ∇EU_{ntw} non-linearly via the CRRA coefficient, and varies across two wave-specific blocks of observations instead of being time-invariant. The unconditional likelihood function $L_n(\theta, \mu)$ does not have a closed-form expression, but can be approximated using simulation methods (Train [2009, p.144-145]). We compute maximum simulated likelihood (MSL) estimates of risk preference parameters θ and the behavioral noise parameter μ by maximizing a simulated analogue to the sample log-likelihood function $\sum_n \ln(L_n(\theta, \mu))$. The estimation sample is 413 subjects who participated in the first experiment or both experiments.

Our modeling framework offers several ways to define and analyze temporal stability of risk attitudes. One can test if the entire population distribution of risk preferences is stable, which can be expressed as a joint hypothesis $H_0: \bar{r}_1 = \bar{r}_2$ and $\sigma_{r1} = \sigma_{r2}$. Alternatively, one can test the temporal stability of the average person’s risk attitude ($H_0: \bar{r}_1 = \bar{r}_2$), or test the temporal stability of unobserved preference heterogeneity ($H_0: \sigma_{r1} = \sigma_{r2}$). We can also accommodate observed heterogeneity by writing $\bar{r}_1$ and $\bar{r}_2$ as linear functions of the subject’s characteristics, such as age, gender and income. It is then possible to consider the question of which demographic groups tend to be more risk averse, and examine if the answer to that question is temporally stable.

The questions so far pertain to temporal stability at the population level, but the analysis can focus on temporal stability at the individual level as well. By normalizing the scale of covariance σ_{r1r2} , one can derive a coefficient $\rho_{r1r2} = \sigma_{r1r2} / (\sigma_{r1} \times \sigma_{r2})$ that directly measures the within-individual correlation of the CRRA coefficients over time. Andersen, Harrison, Lau and Rutström [2008] elicit

risk preferences using multiple price list formats popularized by Holt and Laury [2002], and compute this type of correlation based on the midpoints of CRRA intervals that predict observed behavior under EUT. The approach we take here is far more general because it allows for behavioral errors and can be applied with any elicitation format, as long as the statistical model incorporates a random coefficient specification similar to ours. Moreover, as reported below, one can estimate the within-individual correlations of structural parameters in an analogous manner after correcting for selection and attrition biases and also in the context of RDU models.

B. EUT Specification with Endogenous Sample Selection and Panel Attrition

The experimental design allows us to correct for sample selection into both panel waves of the experiment.¹⁹ Estimates of risk aversion could be sensitive to the sample selection and attrition process in any longitudinal setting, and the estimated coefficients in the behavioral model may be significantly biased if subjects condition their participation on unobservable characteristics that correlate with their latent risk preferences. It is not obvious that individuals with stable preferences are more likely to self-select into the early or later stages of our experiment. Since the decision to participate in the experiment may be correlated with individual risk preferences, it is appropriate to account for possible sample selection and attrition effects in the statistical model.²⁰

To control for sample selection bias, we take the initial pool of 1,996 invited subjects as a random sample from the population, and model the initial selection process that lead to 413 subjects in the first experiment. From this sample of 413, 354 subjects were invited to the second experiment. To control for panel attrition bias, we take those 354 subjects as a random sample from

¹⁹ Vella [1998] surveys alternative specifications for modelling sample selection, including semi-parametric methods.

²⁰ Harrison, Lau and Rutström [2007][2009] use data from a single panel of a previous Danish field experiment that was conducted in June 2003 and correct for sample selection in their analysis of risk attitudes. They use an EUT specification of risk preferences and find evidence of sample selection. The recruited sample of 253 subjects is significantly more risk averse than the general population, but the estimated marginal effects of individual characteristics are similar with and without correction for sample selection.

the sub-population that self-selected into the first experiment, and model the attrition process that led to 182 subjects in the second experiment. This general strategy is consistent with our experimental design, under which the experimenter exogenously determines whether someone is invited to the first experiment, and which subjects in the first experiment get invited to the second experiment.

We first describe a system of binary response models that describes sample selection and attrition. Let s_{nw} be an indicator of whether subject n accepted the invitation to the experiment in wave w ($s_{nw} = 1$) or not ($s_{nw} = 0$). For those who were not invited to the second experiment, we set $s_{n2} = -1$. Assume that each observed outcome s_{nw} is determined by a latent propensity S_{nw} , such that $s_{n1} = \mathbf{I}[S_{n1} > 0]$, and $s_{n2} = \mathbf{I}[S_{n1} > 0 \cap S_{n2} > 0]$ if subject n was invited to the second experiment. The latent propensities are specified as

$$S_{n1} = X_{n1}\beta_1 + u_{n1} = X_{n1}\beta_1 + (a_{n1} + e_{n1}) \quad (7)$$

$$S_{n2} = X_{n2}\beta_2 + u_{n2} = X_{n2}\beta_2 + (a_{n2} + e_{n2}) \quad (8)$$

where X_{nw} is a vector of explanatory variables including a constant, β_w is a conformable vector of coefficients to estimate, and u_{nw} is a random disturbance. We decompose u_{nw} further into a_{nw} and e_{nw} , which are orthogonal to each other. The term a_{nw} captures unobserved characteristics which are potentially correlated with risk attitudes, and across selection and attrition processes. In contrast, e_{nw} captures purely idiosyncratic errors.

Assume that the correlated components a_{n1} and a_{n2} are bivariate normal, and that each idiosyncratic error e_{nw} is independently normal. Under this assumption, the composite errors u_{n1} and u_{n2} are also bivariate normal. When viewed in isolation from the random coefficient EUT model, the system of equations (7) and (8) is analogous to the probit model with sample selection (Van de Ven and Van Praag [1981]) which views the sample retention indicator s_{n2} as the primary outcome of interest.²¹ It is common to normalize this type of model by setting $\text{Var}(u_{n1}) = \text{Var}(u_{n2}) = 1$, and

²¹ The first formal statement of the probit model with sample selection considered the case in which the latent index was the difference in expected utility from two outcomes, which we denote by ∇EU : see Van de Ven and Van Praag [1981; p.235, equation (8)].

identify β_1, β_2 and $\varrho_{s_1s_2} = \text{Corr}(u_{n1}, u_{n2}) = \text{Corr}(a_{n1}, a_{n2})$. We could follow the same convention, but prefer to normalize the system by setting $\text{Var}(u_{n1}) = 2$ and $\text{Var}(u_{n2}) = 2 + \text{Cov}(a_{n1}, a_{n2})$, and identify β_1, β_2 and $\sigma_{s_1s_2} = \text{Cov}(u_{n1}, u_{n2}) = \text{Cov}(a_{n1}, a_{n2})$. This scheme allows us to assume $\text{Var}(a_{n1}) = \text{Var}(e_{n1}) = \text{Var}(e_{n2}) = 1$ and $\text{Var}(a_{n2}) = 1 + \sigma_{s_1s_2}$ without loss of generality; then, equations (7) and (8) can more easily be combined with the random coefficient EUT model by attaching probit probabilities to equation (5), as shown below.

Let $g(a_{n1}, a_{n2}, r_{n1}, r_{n2}; \Theta)$ denote a density function for the joint distribution of risk attitudes and relevant selection/attrition errors, which is characterized by parameters in Θ . Let σ_{s_1rw} and σ_{s_2rw} denote $\text{Cov}(a_{n1}, r_{nw})$ and $\text{Cov}(a_{n2}, r_{nw})$ respectively. We allow for the full set of correlations amongst the four random components. Given the earlier assumptions, $g(\cdot; \Theta)$ is then multivariate normal and $\Theta = (\theta, \Sigma)$, where $\theta = (\bar{r}_1, \bar{r}_2, \sigma_{r1}, \sigma_{r1}, \sigma_{r1r2})$ characterizes the population distribution of the CRRA coefficients and $\Sigma = (\sigma_{s_1s_2}, \sigma_{s_1r1}, \sigma_{s_1r2}, \sigma_{s_2r1}, \sigma_{s_2r2})$ collects covariance parameters that may induce selection and attrition biases. For example, a positive σ_{s_1r1} means that those with relatively large CRRA coefficients in wave 1 are more likely to participate in the first experiment, and a positive σ_{s_2r1} means that such subjects with high CRRA coefficients in wave 1 are also more likely to participate in the second experiment. Without correction for selection and attrition, one would overestimate the initial degree of risk aversion in the population. While $\sigma_{s_1s_2}$ does not address risk attitudes directly, this parameter corrects the attrition process for initial selection bias, since the attrition outcomes are only observed for the self-selected sample of participants in the first experiment. If $\sigma_{s_1s_2}$ is falsely constrained to 0, the resulting correction for attrition bias becomes invalid.

We now turn to a likelihood function which augments the baseline EUT specification with controls for selection and attrition biases. Conditional on a particular set of a_{n1}, a_{n2}, r_{n1} and r_{n2} , the joint likelihood of subject n 's selection/attrition outcomes and risky choices can be specified as

$$\begin{aligned}
\text{CL}_n(a_{n1}, a_{n2}, r_{n1}, r_{n2}, \mu) &= 1 - \Phi(\tau_{n1}) && \text{if } s_{n1} = 0 \\
&= \Phi(\tau_{n1}) \times \prod_t P_{nt1}(r_{n1}, \mu) && \text{if } s_{n1} = 1, s_{n2} = -1 \\
&= \Phi(\tau_{n1}) \times (1 - \Phi(\tau_{n2})) \times \prod_t P_{nt1}(r_{n1}, \mu) && \text{if } s_{n1} = 1, s_{n2} = 0 \\
&= \Phi(\tau_{n1}) \times \Phi(\tau_{n2}) \times \prod_t P_{nt1}(r_{n1}, \mu) \times \prod_t P_{nt2}(r_{n2}, \mu) && \text{if } s_{n1} = 1, s_{n2} = 1
\end{aligned} \tag{9}$$

where $\tau_{nw} = X_{nw}\beta_w + a_{nw}$, $\Phi(\cdot)$ is the standard normal CDF and $P_{nw}(\cdot)$ is the choice-level likelihood under the baseline EUT model. The exact form of the conditional likelihood function thus varies for those who rejected the first invitation ($s_{n1} = 0$), those who participated in the first experiment but did not receive the second invitation ($s_{n1} = 1, s_{n2} = -1$), those who participated in the first experiment but rejected the second invitation ($s_{n1} = 1, s_{n2} = 0$), and finally those who participated in both experiments ($s_{n1} = s_{n2} = 1$). The unconditional likelihood function for subject n can be obtained by taking the expected value of $CL_n(a_{n1}, a_{n2}, r_{n1}, r_{n2}, \mu)$ over the joint distribution of the four random components

$$L_n(\Theta, \mu) = \iiint CL_n(a_{n1}, a_{n2}, r_{n1}, r_{n2}, \mu) g(a_{n1}, a_{n2}, r_{n1}, r_{n2}; \Theta) da_{n1} da_{n2} dr_{n1} dr_{n2}. \quad (10)$$

where $\Theta = (\bar{r}_1, \bar{r}_2, \sigma_{r1}, \sigma_{r2}, \sigma_{r1r2}, \sigma_{s1s2}, \sigma_{s1r1}, \sigma_{s1r2}, \sigma_{s2r1}, \sigma_{s2r2})$ in full. Since equation (10) does not have a closed form expression, we compute the MSL estimates of Θ and μ by maximizing a simulated analogue to the sample log-likelihood function $\sum_n \ln(L_n(\Theta, \mu))$. The estimation sample is all 1,996 subjects who were invited to the first experiment.

Parametric models with selection and attrition such as ours are theoretically identified without the aid of cross-equation exclusion restrictions. Nevertheless, our experimental design provides natural candidates for such restrictions that we use to assist empirical identification. The initial invitation letter randomized subjects to different recruitment fees, and the longitudinal design allows us to observe each subject's additional earnings from the first experiment. Before coming to the first experiment, subjects did not know anything about the 40 lottery pairs used and, during the first experiment, everyone faced the same set of 40 lottery pairs. We assume that the recruitment fees affect the initial decision to accept the first invitation, but do not affect the decision to accept the second invitation once we control for additional earnings from the first experiment.²² We maintain the usual hypothesis that the recruitment fees and prior earnings do not affect the subject's evaluation of lottery pairs directly.

²² Additional earnings in the first experiment include payments for choices in three sets of decision tasks which elicit individual risk attitudes, discount rates and correlation aversion, respectively.

The preceding discussion motivates us to include the recruitment fees only in X_{n1} for the selection equation, the actual earnings from the first experiment only in X_{n2} for the attrition equation, and the lottery payoffs and probabilities only in ∇EU_{nit} for the structural model of risky choices. Our extended specifications condition ∇EU_{nit} on the subject's age, gender and self-reported income via the mean CRRA coefficient $\bar{r}_w$, to capture observed heterogeneity in risk preferences. We include the same set of characteristics in X_{n2} , but only age and gender in X_{n1} since self-reported income is not available for those who rejected the first invitation.

To see the flexibility of our extended specification, one may compare it with several special cases. Consider first a “naïve” approach, in which each panel wave is evaluated separately, using equation (7) to correct for selection into the first wave and equation (8) to correct for selection into the second wave. This approach is naïve in the sense that it fails to recognize the longitudinal nature of the experiments, and requires $\sigma_{s1s2} = \sigma_{s1r2} = \sigma_{s2r1} = 0$. However, even when these restrictions are valid, the approach cannot identify σ_{r1r2} and hence ϱ_{r1r2} that measures the temporal stability of risk preferences within individuals. Two special cases arise if both waves are analyzed jointly, but they correct for only selection bias or attrition bias. With correction for selection bias only, one can estimate all structural parameters consistently when $\sigma_{s2r1} = \sigma_{s2r2} = 0$. The other special case ignores selection bias and requires $\sigma_{s1s2} = \sigma_{s1r1} = \sigma_{s1r2} = 0$. The latter case is perhaps more interesting, considering that it resembles what one would do in typical longitudinal studies that observe no information on those who did not participate in the first wave.

Our modeling strategy provides a general framework for the structural estimation of risk preferences with correction for endogenous selection and attrition. While we parameterize the statistical model using multivariate normal densities and probit kernels, with a few notational changes the likelihood functions above can incorporate other joint distributions of $\{a_{n1}, a_{n2}, r_{n1}, r_{n2}\}$ and kernel CDFs. We focus on the multivariate normal-probit kernel specification primarily to reach a wider audience; the workhorse sample selection models in the empirical literature assume either

the bivariate normality of selection and structural errors or the marginal normality of selection errors.

C. Rank Dependent Utility Theory Specifications

RDU is a popular generalization of EUT, due to Quiggin [1982], that allows the decision-maker to transform the objective probabilities presented in lotteries and use these weighted probabilities to determine decision weights when evaluating lotteries. If $w(p)$ is the probability weighting function assumed, and each lottery has only two prizes such that $M_{i1} > M_{i2}$, then we have

$$RDEU_i = [w(p(M_{i1})) \times U(M_{i1})] + [(1-w(p(M_{i1}))) \times U(M_{i2})], \quad (2')$$

where $RDEU_i$ refers to rank dependent expected utility of lottery i , and the remaining notation is as defined in the context of equation (2).

The logic behind our econometric specifications extends naturally to RDU, once we replace EU_i with $RDEU_i$. Of course, one has to specify the functional form for $w(p)$ and estimate additional parameters. Prelec [1998] offers a two-parameter probability weighting function that exhibits considerable flexibility. This function is

$$w(p) = \exp\{-\eta(-\ln p)^\varphi\}, \quad (12)$$

and is defined for $0 < p < 1$, $\eta > 0$ and $\varphi > 0$. We use its one-parameter special case that assumes $\eta = 1$, and model φ as a log-normally distributed random coefficient φ_{nw} that varies across individuals and panel waves. The resulting one-parameter function exhibits inverse-S probability weighting (optimism for small p , and pessimism for large p) for $\varphi < 1$, S-shaped probability weighting (pessimism for small p , and optimism for large p) for $\varphi > 1$, and linear probability weighting that reduces RDU to EUT when $\varphi = 1$.²³ It rules out the cases of globally concave (optimism for all p) or

²³ The one-parameter Prelec function is similar to another one-parameter function popularized by Tversky and Kahneman [1992]: $w(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma}$, which is inverse-S ($\gamma < 1$) or S-shaped ($\gamma > 1$). When $\varphi=1$ and η is a free parameter instead, equation (12) collapses to the power function $w(p) = p^\eta$; this function can capture either probability optimism ($\eta < 1$) or pessimism ($\eta > 1$), but not both at the same time. There are several versions of the Prelec [1998] function, since several were specified in his Proposition 1 (p.503). We do not use his versions (A) or (B) that constrain φ to be in the unit interval, since that constraint rules out “S-shaped” probability weighting *a priori*, which we view as an unattractive restriction. The one-

globally convex (pessimism for all p) probability weighting *a priori*, and also implies that the fixed point where $w(p) = p$ occurs at $p = 0.368$ for any value of φ . The two-parameter function can admit concave and convex cases, and also inverse-S or S-shaped probability weighting with other fixed points. But allowing for the unrestricted joint distribution of random coefficients and selection/attrition errors leads to several extra parameters, making the use of the two-parameter function less practical for our purposes.²⁴

One implication of the RDU model is that risk preferences are characterized by more than the concavity of the utility function. The risk premium is a complex function of all of the parameters that define the utility function as well as the probability weighting function. Indeed, a concave utility function might be mitigated by probability “optimism” such that the net effect is risk neutrality or even risk loving. We simply have to examine all parameters to characterize risk preferences in the case of RDU: r and φ .²⁵

4. Results

We are interested in testing several hypotheses. First, is the distribution of risk attitudes in the general adult Danish population temporally stable over the one-year period we consider in the

parameter function we use is a special case of version (C) in his Proposition 1.

²⁴ Allowing for the full set of correlations amongst two CRRA coefficients, two probability weighting coefficients, the selection error and the attrition error mean that the RDU specification with the one-parameter Prelec [1998] function already involves at least 13 more parameters to estimate than the EUT specification. The variance-covariance matrix of random parameters r_{n1} , r_{n2} , φ_{n1} , φ_{n2} , a_{n1} and a_{n2} is a 6-by-6 matrix with 15 distinct covariance parameters and 4 identified variance parameters. In comparison, the EUT specification involves 6 covariance parameters and 2 identified variance parameters. One should also estimate the population mean parameters for φ_{n1} and φ_{n2} , and those of r_{n1} and r_{n2} . Of course, the number of extra parameters increases even further when the mean parameters for the probability weighting function are conditioned on observed characteristics. We have also estimated the RDU model with the two-parameter Prelec specification and the results are available upon request. However, under this specification, one cannot easily define temporal stability of the probability weighting function. For example, one cannot identify the average or median person. While it is straightforward to identify the mean and median of each parameter separately, a person with a mean or median value of η does not necessarily have a mean or median value of φ .

²⁵ The EUT model retains some descriptive value, however. The EUT and RDU models explain the overall risk premium, even if they explain it differently. It is sometimes useful to focus on the parameter r in the EUT model as a summary statistic on the overall risk premium, even if the RDU model may provide the correct structural decomposition into aversion to outcome variability (the r parameter) and probability weighting (the φ parameter).

experiment? Second, are risk attitudes temporally stable at the individual level? Third, does the possibility of non-random sample selection and attrition change our inferences about the temporal stability of risk attitudes?

We use maximum simulated likelihood to estimate the full statistical model that captures unobserved preference heterogeneity, endogenous selection into the first experiment, and endogenous panel attrition between the two experiments. Train [2009] provides details on MSL estimation of heterogeneous preference models without selection. Cappellari and Jenkins [2004] show how one can control for endogenous selection and attrition using MSL in the context of models without unobserved preference heterogeneity. By modeling the joint likelihood of observing the entire series of responses by each subject and adjusting standard errors for clustering at the subject level, our statistical specification allows for “clustered” responses by the same subject. Panel-robust Wald statistics are used to test various hypotheses with respect to the estimated coefficients. The statistical model also allows for heteroscedasticity in the behavioral error term, by conditioning the Fechner noise parameter on binary variables for each treatment in the experimental design; one variable captures the order of risk aversion and discounting tasks, and the other variable captures our use of high and low stakes in the risk aversion tasks. We also condition the population mean coefficients of latent risk preference parameters on these two treatment variables.

We transform several estimates into alternative forms that are easier to interpret. The tables below report correlation coefficients instead of covariance parameters. In case of the log-normal random coefficient φ in the RDU model, all results are for φ itself instead of $\ln(\varphi)$.²⁶ Finally, we divide selection and attrition equation coefficients by the normalized standard deviation of each equation so that they can be interpreted in the same manner as familiar probit coefficients.

²⁶ Specifically, we report the mean of φ for the base group (constant), along with the marginal effect of each observed characteristic on the mean of φ for the base group. The standard deviation of φ is evaluated at the sample average characteristics. The within-individual correlation of φ is computed by applying the usual formula for the correlation coefficient of bivariate log-normal random variables. Other correlations involving φ present cases where we compute the correlation between a log-normal random variable and a normal random variable. Garvey, Book and Covert [2015, p. 443, Theorem B.1] provide a closed-form formula that can be applied to these cases.

A. Temporal Stability of Risk Attitudes

We find mixed evidence of temporal stability for inferred risk attitudes under EUT when the model fully corrects for endogenous sample selection and attrition bias. Table 2 contains these results, including single hypothesis tests that the mean CRRA parameter $\bar{r}_w$ for each treatment group is the same over time. For example, the estimated mean coefficient of relative risk aversion in our baseline treatment is equal to 0.378 in wave 1, and equal to 0.678 in wave 2; the estimated difference in the two mean population coefficients is equal to 0.300, which is significantly different from 0 with a p -value of 0.017. We also find that the estimated population mean coefficient is larger in wave 2 relative to wave 1 when we control for the high stakes treatment; the estimated difference between the two coefficients is 0.276, which is significantly different from 0 with a p -value of 0.024. However, we find that the estimated population standard deviation of relative risk aversion is temporally stable; the estimated standard deviation of the r parameter, σ_r , drops marginally from 0.726 in wave 1 to 0.705 in wave 2, and the estimated difference between the two coefficients is not significantly different (p -value of 0.754). A joint test of estimated mean population coefficients and standard deviations across the two waves allows us to evaluate whether the entire population distribution is temporally stable. The $\chi^2(4)$ test statistic has a p -value of 0.110, so we cannot reject the hypothesis of temporal stability at the 10% significance level.²⁷ Although the estimated population mean is significantly higher in wave 2 compared to wave 1 for low and high stakes treatments, we find statistical evidence of temporal stability for the entire population distribution of relative risk aversion.

The upper panel in Figure 1 shows the estimated population distributions of relative risk aversion across the two waves and two monetary treatments, with controls for non-random selection and attrition bias. We observe very small differences in estimated populated distributions

²⁷ Since the mean of the r parameter has been conditioned on two treatment variables, in each wave there are 3 estimates associated with the mean (constant, RAfirst, RAhigh). Temporal stability of the population distribution therefore entails 4 between-wave equality restrictions, comprising 3 restrictions on the mean and 1 restriction on the standard deviation.

across the two monetary treatments in both waves, but the population distributions of relative risk aversion for both monetary treatments shift to the right in wave 2 compared to wave 1, and illustrate the statistical test results reported above.

We also consider temporal stability at the individual level. The estimated correlation coefficient between relative risk aversion in wave 1 and 2, ρ_{r1r2} , is equal to 0.410, which is significantly different from 0 (p -value of <0.001). The significant positive correlation suggests that risk preferences are temporally stable at the individual level, in the sense that someone with an above-average r parameter in wave 1 also tends to have an above-average r parameter in wave 2, and thus we reject the hypothesis that the two population distributions are independent.

Turning to the results for RDU in Table 3, we draw similar mixed conclusions that depend on which aspect of temporal stability that one is interested in. Under RDU, risk preferences are characterized by the r parameter as well as the weighting parameter, φ , which is log-normally distributed. The entire population distribution of risk preferences may be said to be stable when the joint distribution of r and φ is stable. More formally, this joint hypothesis requires stability in the estimated the population means of the r and φ parameters, the estimated population standard deviations of r and φ , and the estimated correlation between r and φ . We reject this type of temporal stability; the associated $\chi^2(9)$ test statistic has a p -value less than 0.001.²⁸

Figure 2 displays the estimated population distributions of relative risk aversion for each wave and monetary treatment. The estimated distributions in the upper panel control for selection and attrition bias, and we observe the same pattern as before: the population distributions of the r parameter for both monetary treatments shift to the right in wave 2 compared to wave 1, and we observe small marginal effects of the monetary treatments on elicited risk attitudes. We also observe that the population distributions in wave 2 have a smaller standard deviation than the distributions

²⁸ For the same reasons as discussed in footnote 27, the stable marginal distribution of the r parameter entails 4 restrictions. Similarly, the stable marginal distribution of the φ parameter entails another set of 4 restrictions. In total, temporal stability in the joint distribution of r and φ parameters entails 9 between-wave equality restrictions: 8 restrictions on the marginal distributions and 1 restriction on the correlation coefficient between the two parameters.

in wave 1; the estimated standard deviation is 0.612 in wave 1 and 0.491 in wave 2, and we reject the null hypothesis that the estimated difference in the two coefficients is equal to 0 at the 10% significance level (p -value of 0.058). The estimated correlation coefficient between the population distributions of the r parameter over time, ρ_{r1r2} , is equal to 0.437, which is slightly higher than the estimated coefficient under EUT, and we cannot reject the hypothesis that the two distributions are dependent.

The estimated population distributions of the probability weighting parameter φ are displayed in Figure 3. The distributions in the upper panel control for selection and attrition bias, and we again observe small marginal effects of the monetary treatment on the estimated population distributions in each panel. However, we do observe differences in the estimated population distributions of the φ parameter between the two waves, and we reject the hypothesis that the population distribution of the φ parameter is temporally stable at the 10% significance level; the $\chi^2(4)$ test statistic has a p -value of 0.051. We cannot reject the hypothesis that the estimated population means are different between the two waves (the $\chi^2(3)$ test statistic has a p -value of 0.384), but we find that the estimated standard deviation is significantly higher in wave 2 compared to wave 1 (p -value = 0.019) which generates the instability over time. Despite these differences in the estimated population distributions, we find that the estimated between-wave correlation of the φ parameter is 0.852 with a standard error of 0.055, which suggests that there is a strong degree of temporal stability at the individual level.

B. Effects of Sample Selection and Attrition on Risk Attitudes under EUT

We observe significant evidence of exogenous and endogenous selection and attrition effects on the estimated coefficients reported in Table 2. We find a positive significant effect of the higher recruitment fee on the propensity to select into the first wave of our experiment. In effect, the law of demand applies to participation in the experiments, and response rates increase significantly when the recruitment fee is raised from 300 kroner to 500 kroner for participation in wave 1. Middle aged

and older subjects are also more likely to select into the first wave compared to omitted age group. However, it is generally difficult to explain panel retention rates in terms of observed characteristics, although the results do suggest that young subjects and those with high income are less likely to select into the second wave than otherwise.

Turning to endogenous effects of sample selection and attrition, we find enough statistical evidence to reject the hypothesis of no selection and attrition bias, respectively. The hypothesis of no endogenous sample selection bias is evaluated using the joint test of $H_0: \rho_{s1s2} = \rho_{s1r1} = \rho_{s1r2} = 0$. This hypothesis is rejected, with a p -value less than 0.001. The hypothesis of no endogenous attrition bias can be tested by $H_0: \rho_{s2r1} = \rho_{s2r2} = 0$, which again is rejected, with a p -value less than 0.001. The estimated correlation coefficient between the error terms in the selection and attrition equations, ρ_{s1s2} , is equal to -0.533 with a standard error of 0.079, which means that one cannot take the naïve approach of correcting for each source of sampling bias separately.

We can see the overall effects of controlling for selection and attrition bias on the estimated population distributions of relative risk aversion in Figure 1. The results suggests that neglected selection and attrition biases may lead one to draw opposite conclusions about the source of instability. The lower panel shows the estimated distributions with no correction for sample selection and attrition bias, and we observe that over time the distribution becomes tighter around the almost invariant average.²⁹ In contrast, the corrected estimates in the upper panel suggests that there is an increase in the average relative risk aversion over time, and that the spread of the distribution around the average remains more or less stable.

C. Effects of Sample Selection and Attrition on Risk Attitudes under RDU

We observe similar exogenous sample selection and attrition effects under RDU compared to EUT, and continue to observe significant selection and attrition bias under RDU. The hypothesis

²⁹ Table B1 in Appendix B reports the estimated parameters for the EUT model with no correction for selection and attrition bias.

test of no sample selection bias now involves the correlation coefficients between the error term in the selection equation and the five other random components (the error term in the attrition equation, two r parameters, and two φ parameters). This hypothesis is rejected at all conventional levels, since the p -value is less than 0.001. The hypothesis test of no attrition bias involves the correlation coefficients between the error term in the attrition equation and four structural parameters (two r parameters, and two φ parameters) and we again reject the null hypothesis of no attrition bias (the p -value is less than 0.001). The estimated correlation coefficient between the error terms in the selection and attrition equations, ρ_{s1s2} , is equal to -0.236 with a standard error of 0.089, and we can again reject the naïve approach of correcting for each source of sampling bias separately.

Figure 3 displays the overall effects of controlling for selection and attrition bias on the estimated population distributions of the probability weighting parameter. The lower panel shows the estimated distributions with no correction for sample selection and attrition bias, and here we find statistical evidence of temporal stability.³⁰ More specifically, without corrections for non-random selection and attrition bias, we cannot reject the null hypothesis that the population distribution of the φ parameter is temporally stable (the $\chi^2(4)$ test statistic has a p -value of 0.100). Viewed another way, the uncorrected estimates of the probability weighting parameter seem relatively stable around biased base levels, whereas the corrected estimates point to some statistical evidence of temporal instability. We also observe that the shape of the population distribution for the weighting parameter changes when we correct for selection and attrition bias. Figure 3 shows that the population distribution of the φ parameter is more skewed to the right in the upper panel with corrections compared to the lower panel without corrections. A larger fraction of subjects can be classified by an inverse-S shaped probability weighting function when we correct for selection and attrition bias compared to the non-corrected estimates.

We can look closer at the effect of adding controls for sample selection and attrition on risk attitudes under RDU. The effects on the mean of the r parameter are modest: estimates of concavity

³⁰ The estimated parameters are reported in Table B2 in Appendix B.

slightly decline in both wave 1 and wave 2 when we control for selection and attrition bias, so the risk premium derived from utility concavity, *ceteris paribus*, is lower. The effect on the mean of the φ parameter is more striking, and shown in Figure 4. The top 4 panels of Figure 4 refer to the estimates with no corrections for sample selection and attrition, and the bottom 4 panels refer to the estimates with corrections for sample selection and attrition. The panels on the left are, of course, the implied probability weighting functions. The panels on the right are designed to allow one to see the implied decision weights, which are what is important in the end. Consider the decision weights shown in the top right panel. Each line shows the implied decision weights for an equi-probable lottery. One is for a 2-prize lottery with probabilities $\frac{1}{2}$ and $\frac{1}{2}$, one is for a 3-prize lottery with probabilities $\frac{1}{3}$, $\frac{1}{3}$ and $\frac{1}{3}$, and one is for a 4-prize lottery with probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$ and $\frac{1}{4}$. These examples are deliberately using uniform probabilities of outcomes so that one sees the pure effect of probability weighting.

Based on Figure 4, we can infer the effect of probability weighting on risk attitudes for the average Dane. In the top panel we have no corrections and the estimates for wave 1. The “S-shape” of the probability weighting function leads to a *negative* risk premium for the 2-prize lottery, *ceteris paribus*, since the decision weight on the worst outcome is lower than $\frac{1}{2}$ and the decision weight on the better outcome is greater than $\frac{1}{2}$. For 3-prize or 4-prize lotteries the effect is to lead to under-weighting of extreme outcomes and over-weighting of interior outcomes. In wave 2 we see even greater probability weighting with no corrections, also consistent with “S-shaped” probability weighting. The effect of corrections on inferences are striking: the evidence for probability weighting virtually disappears in wave 1 and is dramatically reduced in wave 2. The general lesson here is that one might have concluded that there was significant evidence of strong probability weighting without correcting for sample selection and attrition bias, when in fact the evidence for probability weighting is very slight.

D. Incorporating Observed Heterogeneity

We have estimated the population distributions of structural parameters to account for interpersonal heterogeneity in risk preferences. An alternative way to capture preference heterogeneity is to generalize representative agent models by allowing structural parameters to vary with observed personal characteristics. This type of observed heterogeneity can be incorporated into our analysis by conditioning the population mean of each parameter on the decision maker's characteristics, in the same manner as we have conditioned the mean of each parameter on the treatment variables.

For illustration, we replace the two treatment variables with a female dummy and estimate models that focus on the overall male-female differences in risk preferences. Despite the common assertion that women are more risk averse than men, the supporting evidence is not ubiquitous and previous studies in Denmark do not find significant male-female differences in risk attitudes (Harrison, Lau and Rutström [2007; p.361]). Figure 5 displays the estimated population distributions of the r and φ parameters under the RDU model with correction for selection and attrition biases. In either wave, we do not observe any significant male-female difference in the population mean parameter, both in terms of the utility function and the probability weighting function.³¹ We draw qualitatively similar conclusions about temporal stability for both men and women: there is a significant between-wave change in the mean of the r parameter (p -values of <0.001 for men and 0.048 for women) but not in the mean of the φ parameter (p -values of 0.657 and 0.774 , respectively). The hypotheses of no selection bias and no attrition bias are rejected at the 1% level. Without correction for selection and attrition biases, we would have found a significant between-wave change in the mean of the φ parameter for men but not for women (p -values of 0.081 and 0.148).³²

³¹ The male-female difference in the mean of the r parameter is 0.033 (p -value = 0.740) in wave 1 and 0.102 (p -value = 0.316) in wave 2. The male-female difference in the mean of the φ parameter is 0.284 (p -value = 0.220) in wave 1 and 0.420 (p -value = 0.374) in wave 2.

³² Detailed estimation results for the preceding discussion, as well as parallel results under EUT, are available upon request.

5. Conclusions

Heckman and Smith [1995; p.99] noted that, “Surprisingly, little is known about the empirical importance of randomization bias.” Aggregate data on participation rates from job training experiments by Hotz [1992] and clinical trials by Kramer and Shapiro [1984] suggest that the bias due to endogenous participation decisions might be significant, but we know of no study that directly evaluates the hypothesis.³³ We do not *a priori* know the direction of randomization bias in economics experiments, and whether the use of recruitment fees may mitigate the effects of randomization bias on elicited risk attitudes.

Our results suggest that randomization bias can have significant effects on inferences about risk attitudes. Neglecting endogenous sample selection and attrition may lead one to draw erroneous conclusions about risk attitudes at a point in time (e.g. the average Dane’s relative risk aversion now), as well as stability in risk attitudes over time (e.g. whether the average Dane’s relative risk aversion has changed over time). These conclusions hold whether one uses an EUT or RDU characterization of risk attitudes, although the way in which sample selection and attrition affects the analysis is different across the two decision theories as well as alternative measures of temporal stability that one may consider.

These effects of randomization bias on risk attitudes are clear in our design because of the exogenous variation in recruitment fees. We do not claim that our findings generalize beyond the adult Danish population, the specific recruitment fees we employed, or the battery of lotteries we employed. On the other hand, our sample is wide and representative of the adult Danish population, and our recruitment fees and lottery parameters fall well within common practice in field

³³ Many other hypotheses about the effects of sample selection and attrition in longitudinal studies have been evaluated, of course. In the case of clinical trials, for instance, Beunckens, Molenberghs and Kenward [2005] compare the effects of obvious *ad hoc* methods (such as assuming that the last observed case for some subject who does not participate in later sessions is the observation that the subject would have provided, or only using sub-samples that participate in all sessions), methods based on imputation and corrections for the imprecision of the imputation, and “direct-likelihood” methods such as those used here.

experiments. The constructive implication for future experimental design is to exogenously vary show-up fees and evaluate the effects on a case-by-case basis.

If the need for corrections to mitigate randomization bias is “bad news” from our results, the “good news” is that even after making such corrections, there are still many quantitative and qualitative aspects of risk attitudes that remain temporal stable, at least for this population and the time frame evaluated in our experiments.

Table 1: Sample Sizes and Descriptive Statistics*A. Sample Sizes*

Recruitment	Variable	Wave 1	Wave 2	All
High Fixed Fee	Invited	865	184	1049
	Accepted	208	89	297
	Percent Accept	24.1%	48.4%	28.3%
Low Fixed Fee	Invited	1131	170	1301
	Accepted	205	93	298
	Percent Accept	18.1%	54.7%	22.9%

B. Descriptive Statistics for Participants

Variable	Definition	Mean Wave 1	Mean Wave 2
female	Female	0.48	0.45
young	Aged less than 30	0.16	0.13
middle	Aged between 40 and 50	0.23	0.21
old	Aged over 50	0.49	0.53
IncLow	Lower level income	0.22	0.23
IncHigh	Higher level income	0.47	0.45
Number of subjects		413	182

Notes: Most variables have self-evident definitions. The omitted age group is 30-39. Lower income is defined in variable “IncLow” by a household income in 2008 below 300,000 kroner. Higher incomes are defined in variable “IncHigh” by a household income of 500,000 kroner or more.

**Table 2: Estimates of EUT Parameters
with Full Controls for Sample Selection and Attrition**

(Log-simulated likelihood = -10806 for 25,555 observations on 413 subjects and 1,583 rejecters in wave 1 and 182 subjects and 172 rejecters in wave 2 using 100 Halton draws.)

Variable	Estimate	Standard Error	p -value	95% Confidence Interval	
<i>Selection equation: $\beta_1/\sqrt{Var(u_{n1})}$</i>					
female	-0.096	0.062	0.119	-0.218	0.025
young	0.212	0.117	0.069	-0.017	0.441
middle	0.239	0.107	0.026	0.028	0.449
old	0.272	0.097	0.005	0.082	0.461
high_fee	0.186	0.066	0.004	0.058	0.315
constant	-1.054	0.098	0.000	-1.247	-0.862
<i>Attrition equation: $\beta_2/\sqrt{Var(u_{n2})}$</i>					
female	-0.081	0.105	0.437	-0.286	0.124
young	-0.364	0.201	0.071	-0.758	0.031
middle	-0.092	0.172	0.595	-0.429	0.246
old	0.005	0.152	0.975	-0.294	0.303
IncLow	-0.146	0.133	0.274	-0.406	0.115
IncHigh	-0.238	0.115	0.038	-0.464	-0.013
earnings	0.011	0.029	0.694	-0.045	0.068
constant	0.979	0.182	0.000	0.623	1.336
<i>Mean of r parameter in wave 1</i>					
RAfirst	0.097	0.083	0.245	-0.066	0.260
RAhigh	0.066	0.025	0.008	0.017	0.115
constant	0.378	0.074	0.000	0.233	0.524
<i>Mean of r parameter in wave 2</i>					
RAfirst	0.034	0.059	0.563	-0.082	0.150
RAhigh	0.043	0.036	0.234	-0.028	0.113
constant	0.678	0.088	0.000	0.506	0.850

Standard deviations and correlation coefficient of r parameters in wave 1 and wave 2

σ_{r1}	0.726	0.058	0.000	0.612	0.840
σ_{r2}	0.705	0.065	0.000	0.578	0.832
ϱ_{r1r2}	0.410	0.115	0.000	0.185	0.636

Other correlation coefficients

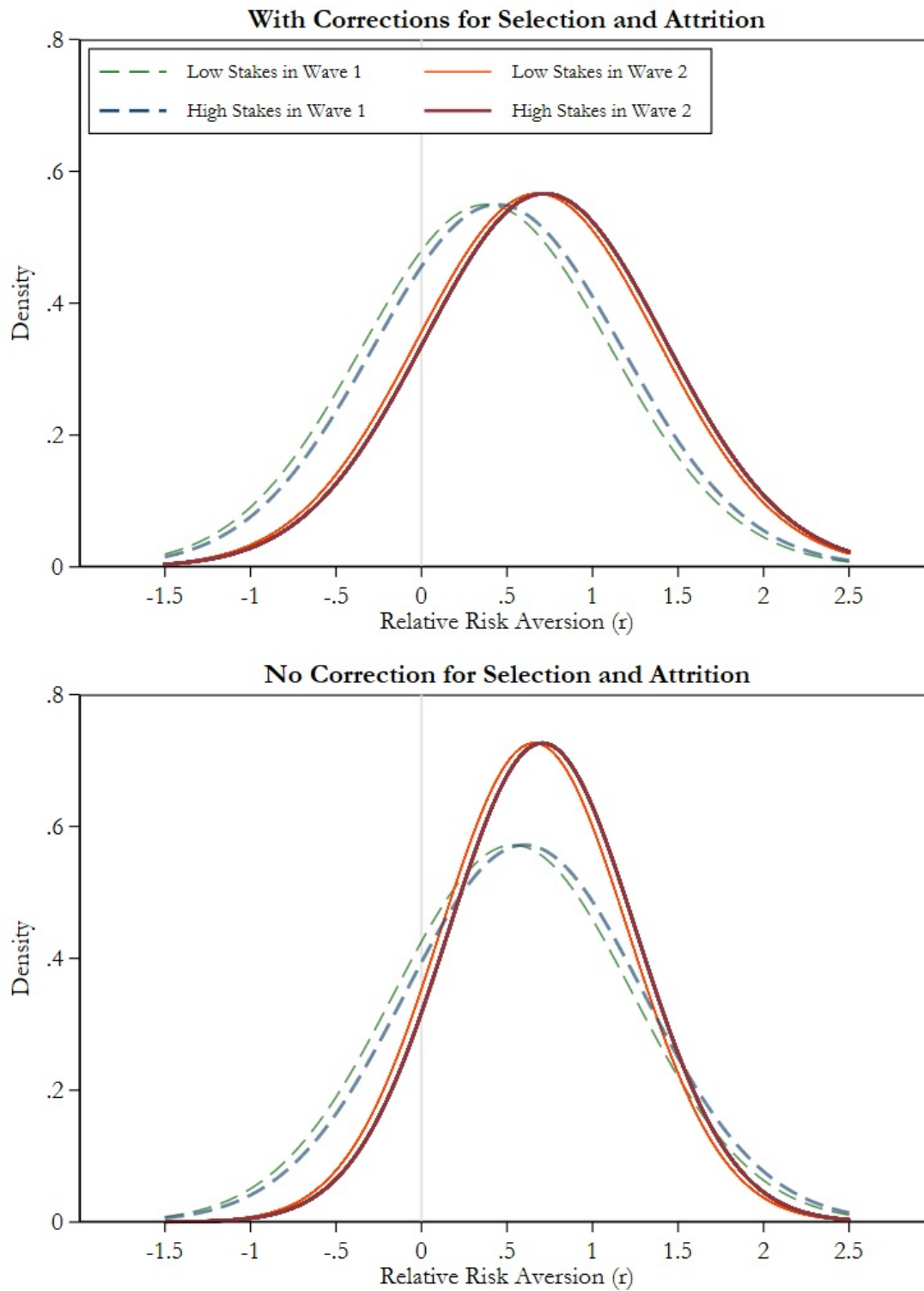
ϱ_{s1s2}	-0.533	0.079	0.000	-0.688	-0.379
ϱ_{s1r1}	0.062	0.062	0.319	-0.060	0.183
ϱ_{s1r2}	-0.500	0.050	0.000	-0.598	-0.402
ϱ_{s2r1}	-0.120	0.075	0.109	-0.267	0.027
ϱ_{s2r2}	0.748	0.057	0.000	0.635	0.860

Test for temporal stability of predicted group means for r parameter

ΔBase	0.300	0.125	0.017	0.054	0.545
ΔRAhigh	0.276	0.123	0.024	0.036	0.517
$\Delta\text{RAfirst}$	0.237	0.153	0.122	-0.063	0.537

Notes: Group means are predicted using the estimated mean function for r parameter. ΔBase tests whether the between-wave difference in constant is significant. ΔRAhigh ($\Delta\text{RAfirst}$) tests whether the between-wave difference in constant + RAhigh (RAfirst) is significant.

Figure 1: Population Distributions of Risk Aversion under EUT



**Table 3: Estimates of the RDU Parameters
with Full Controls for Sample Selection and Attrition**

(Log-simulated likelihood = -9724 for for 25,555 observations on 413 subjects and 1,583 rejecters in wave 1 and 182 subjects and 172 rejecters in wave 2 using 100 Halton draws.)

Variable	Estimate	Standard Error	<i>p</i> -value	95% Confidence Interval	
<i>Selection equation: $\beta_1/\sqrt{Var(u_{n1})}$</i>					
female	-0.059	0.061	0.332	-0.180	0.061
young	0.164	0.109	0.133	-0.050	0.379
middle	0.268	0.102	0.008	0.069	0.468
old	0.329	0.092	0.000	0.150	0.509
high_fee	0.192	0.061	0.002	0.073	0.313
constant	-1.072	0.091	0.000	-1.251	-0.894
<i>Attrition equation: $\beta_2/\sqrt{Var(u_{n2})}$</i>					
female	-0.113	0.128	0.380	-0.364	0.139
young	-0.321	0.257	0.211	-0.825	0.183
middle	-0.206	0.237	0.386	-0.671	0.259
old	-0.068	0.206	0.742	-0.472	0.337
IncLow	-0.256	0.183	0.160	-0.615	0.102
IncHigh	-0.165	0.157	0.295	-0.473	0.144
earnings	0.027	0.043	0.533	-0.057	0.111
constant	0.682	0.251	0.007	0.190	1.174
<i>Mean of <i>r</i> parameter in wave 1</i>					
RAfirst	0.092	0.082	0.263	-0.069	0.253
RAhigh	0.056	0.034	0.097	-0.010	0.122
constant	0.356	0.074	0.000	0.212	0.501
<i>Mean of <i>r</i> parameter in wave 2</i>					
RAfirst	-0.164	0.072	0.023	-0.305	-0.023
RAhigh	0.023	0.048	0.635	-0.072	0.118
constant	0.548	0.077	0.000	0.397	0.700

Standard deviations and correlation coefficient of r parameters in wave 1 and wave 2

σ_{r1}	0.612	0.062	0.000	0.491	0.733
σ_{r2}	0.491	0.044	0.000	0.405	0.576
ϱ_{r1r2}	0.437	0.076	0.000	0.289	0.585

Mean of φ parameter in wave 1

RAfirst	-0.015	0.121	0.903	-0.251	0.222
RAhigh	0.061	0.061	0.316	-0.059	0.181
constant	1.057	0.153	0.000	0.757	1.357

Mean of φ parameter in wave 2

RAfirst	0.254	0.191	0.183	-0.120	0.628
RAhigh	0.008	0.097	0.983	-0.183	0.198
constant	1.247	0.206	0.000	0.842	1.651

Standard deviations and correlation coefficient of φ parameters in wave 1 and wave 2

$\sigma_{\varphi1}$	2.118	0.480	0.000	1.177	3.060
$\sigma_{\varphi2}$	4.428	1.536	0.004	1.417	7.439
$\varrho_{\varphi1\varphi2}$	0.852	0.055	0.000	0.744	0.959

Other correlation coefficients

ϱ_{s1s2}	-0.236	0.089	0.008	-0.411	-0.062
ϱ_{s1r1}	0.184	0.079	0.021	0.028	0.339
ϱ_{s1r2}	0.200	0.090	0.025	0.025	0.376
$\varrho_{s1\varphi1}$	0.384	0.030	0.000	0.326	0.443
$\varrho_{s1\varphi2}$	0.306	0.045	0.000	0.219	0.394
ϱ_{s2r1}	-0.346	0.051	0.000	-0.446	-0.247
ϱ_{s2r2}	0.096	0.085	0.258	-0.071	0.264
$\varrho_{s2\varphi1}$	-0.245	0.076	0.001	-0.394	-0.097
$\varrho_{s2\varphi2}$	-0.153	0.067	0.022	-0.284	-0.023
$\varrho_{r1\varphi1}$	0.086	0.043	0.044	0.002	0.170
$\varrho_{r1\varphi2}$	0.075	0.056	0.183	-0.035	0.185
$\varrho_{r2\varphi1}$	0.269	0.044	0.000	0.182	0.356
$\varrho_{r2\varphi2}$	0.178	0.044	0.000	0.092	0.264

Test for temporal stability of predicted group means for r parameter

Δ Base	0.192	0.118	0.105	-0.040	0.424
Δ RAhigh	0.159	0.102	0.120	-0.041	0.359
Δ RAfirst	-0.064	0.124	0.608	-0.308	0.180

Test for temporal stability of predicted group means for φ parameter

Δ Base	0.190	0.207	0.359	-0.215	0.595
Δ RAhigh	0.136	0.193	0.481	-0.242	0.514
Δ RAfirst	0.458	0.196	0.019	0.074	0.843

Notes: Group means are predicted using the estimated mean function for each parameter. Δ Base tests whether the between-wave difference in constant is significant. Δ RAhigh (Δ RAfirst) tests whether the between-wave difference in constant + RAhigh (RAfirst) is significant.

**Figure 2: Population Distributions of
Risk Aversion Due to Utility
Curvature under RDU**

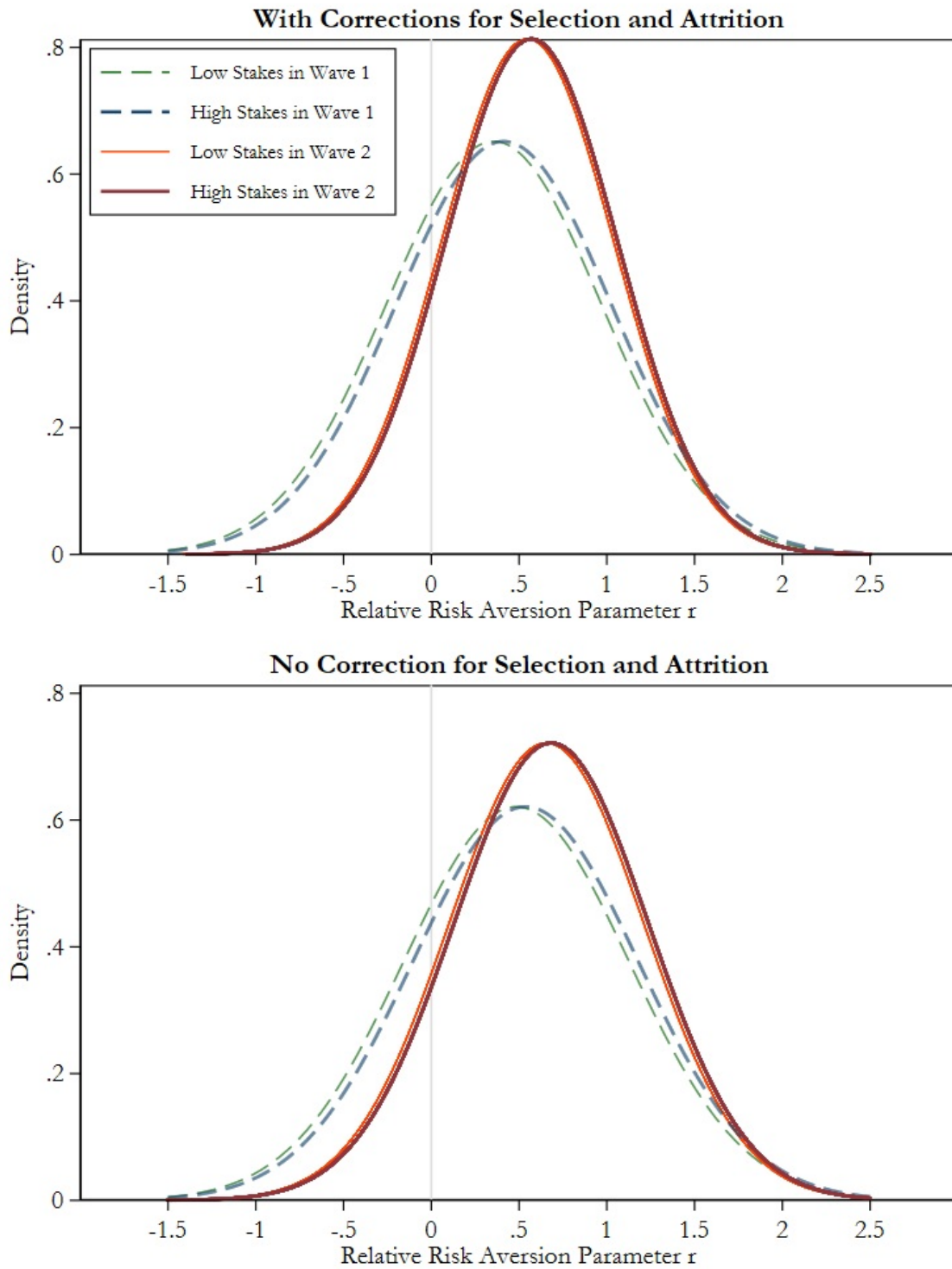


Figure 3: Population Distributions of Risk Aversion Due to Probability Weighting under RDU

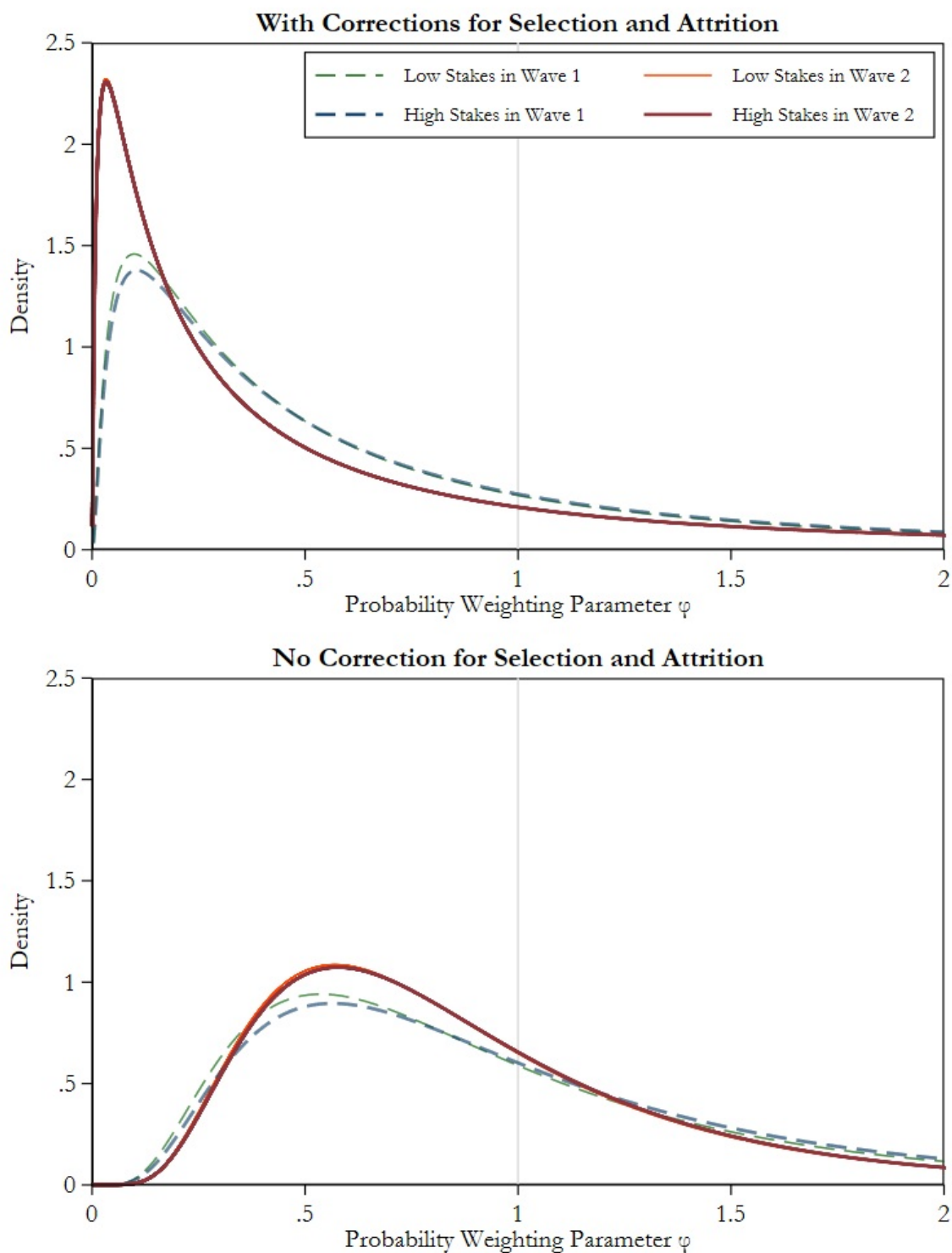


Figure 4: Probability Weighting, Decision Weights, and Corrections for Sample Selection and Attrition

Without SS corrections: $r = .48$ (wave 1) and $.66$ (wave 2)

With SS corrections: $r = .36$ (wave 1) and $.55$ (wave 2)

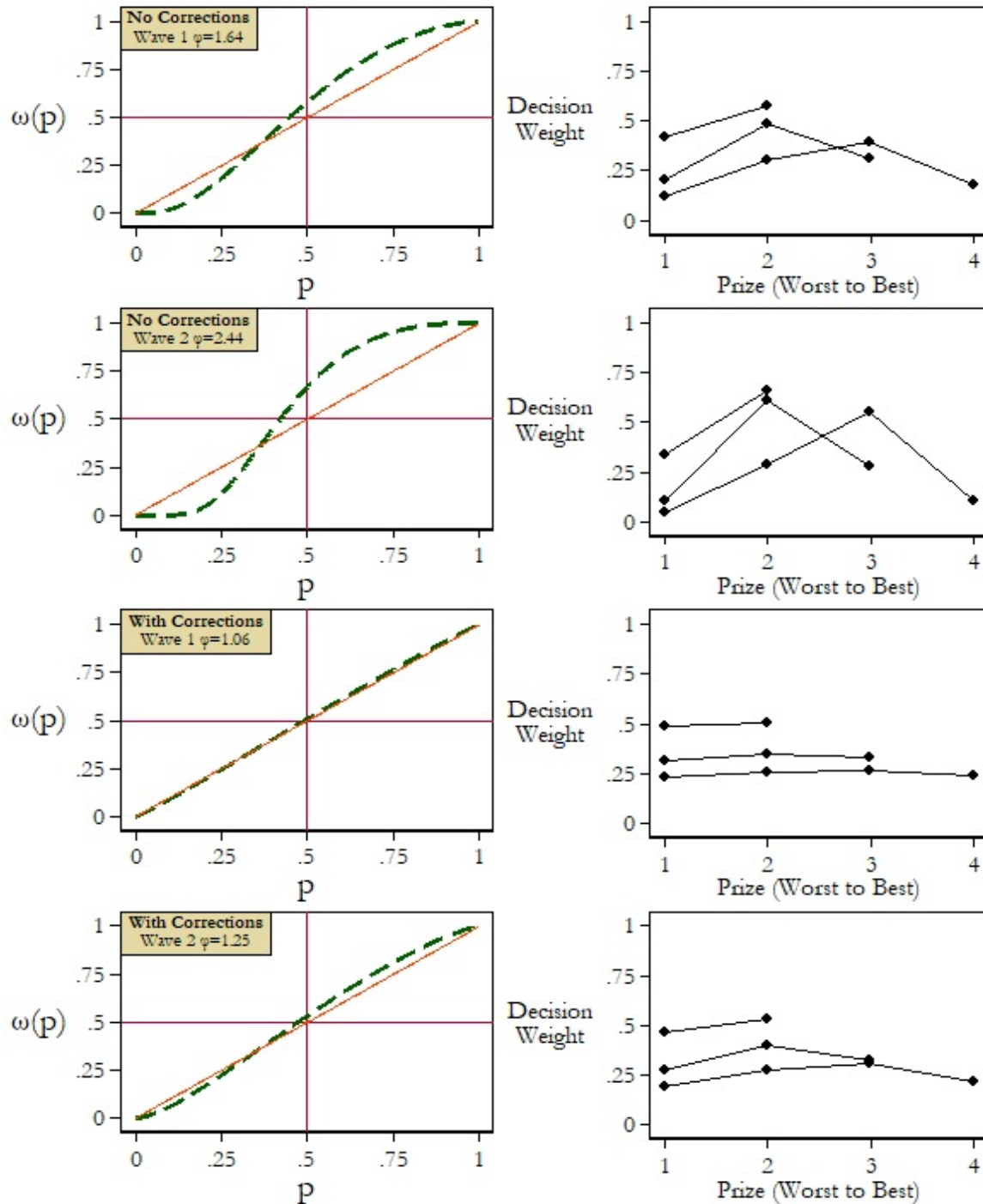
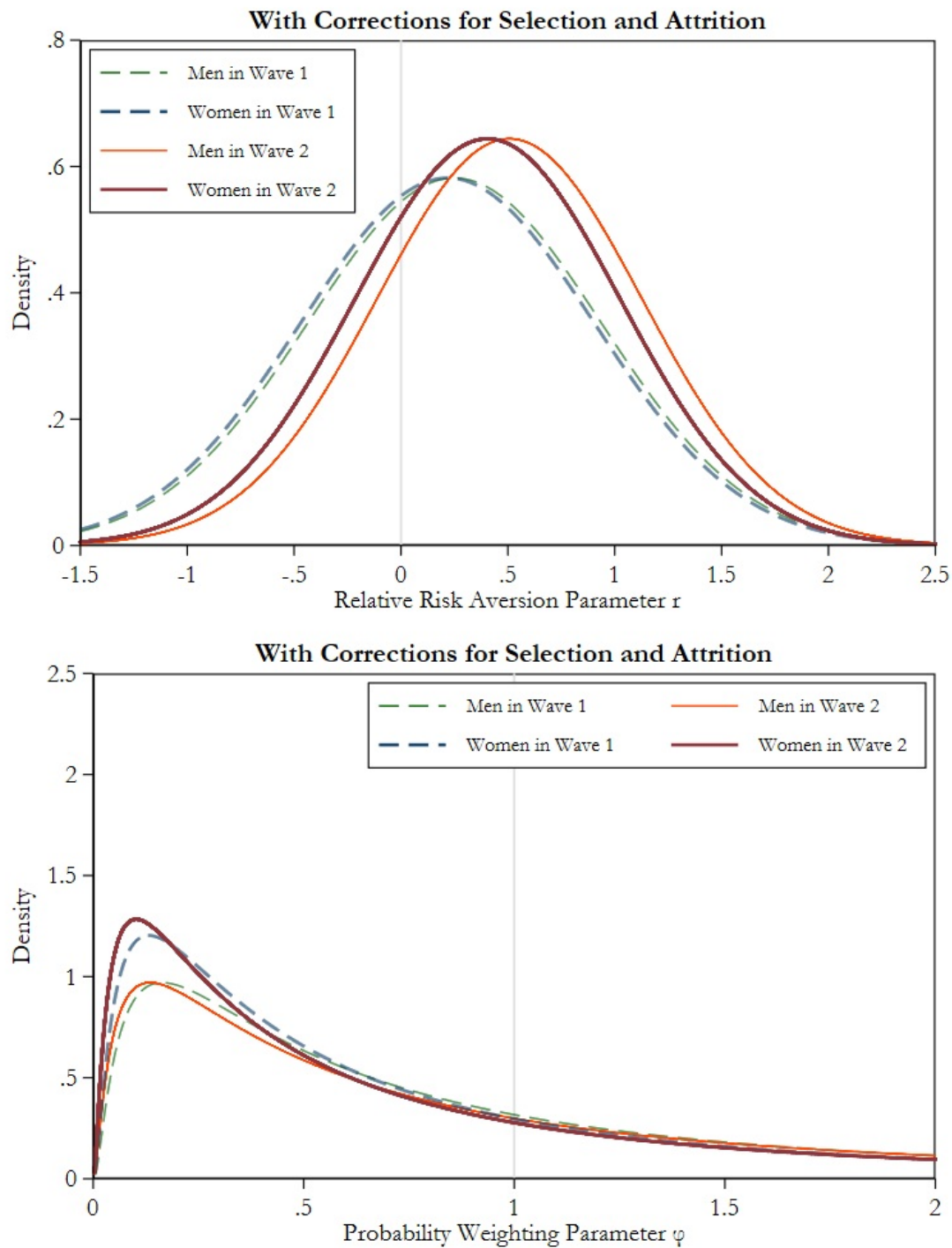


Figure 5: Population Distributions of Risk Aversion for Men and Women under RDU



References

- Andersen, Steffen; Cox, James C.; Harrison, Glenn W.; Lau, Morten; Rutström, E. Elisabet, and Sadiraj, Vjollca, "Asset Integration and Attitudes to Risk: Theory and Evidence," *Working Paper 2011-17*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2011.
- Andersen, Steffen, Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, "Lost in State Space: Are Preferences Stable?" *International Economic Review*, 49(3), 2008, 1091-1112.
- Andersen, Steffen, Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, "Discounting Behavior: A Reconsideration" *European Economic Review*, 71(1), 2014, 15-33.
- Andersen, Steffen, Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, "Multiattribute Utility Theory, Intertemporal Utility and Correlation Aversion?" *International Economic Review*, 2017, forthcoming.
- Baucells, Manel, and Villasís, Antonio, "Stability of risk preferences and the reflection effect of prospect theory," *Theory and Decision*, 68, 2010, 193-211.
- Behaghel, Luc; Crépon, Bruno; Gurgand, Marc, and Le Barbanchon, Thomas, "Sample Attrition Bias in Randomized Experiments: A Tale of Two Surveys," *IZA Discussion Paper #4162*, Institute for the Study of Labor (IZA), Bonn, May 2009.
- Beunckens, Caroline; Molenberghs, Geert, and Kenward, Michael G., "Direct Likelihood Analysis Versus Simple Forms of Imputation for Missing Data in Randomized Clinical Trials," *Clinical Trials*, 2, 2005, 379-386.
- Capellari, Lorenzo, and Jenkins, Stephen P., "Modelling Low Income Transitions," *Journal of Applied Econometrics*, 19(5), 2004, 593-610.
- Cleave, Blair L.; Nikiforakis, Nikos, and Slonim, Robert, "Is There Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences," *Experimental Economics*, 16, 2013, 372-382.
- Das, Mitali; Newey, Whitney, and Vella, Francis, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70(1), January 2003, 33-58.
- Dasgupta, Utteeyo; Gangadharan, Lata; Maitra, Pushkar, and Mani, Subha, "Searching for preference stability in a state dependent world," *Journal of Economic Psychology*, 62, 2017, 17-32.
- Demirtas, Hakan, and Schafer, Joseph L., "On the performance of random-coefficient pattern-mixture models for non-ignorable dropout," *Statistics in Medicine*, 22, 2003, 2553-2575.
- Diggle, P., and Kenward, M.G., "Informative Drop-out in Longitudinal Data Analysis," *Applied Statistics*, 43(1), 1994, 49-93.

- DiNardo, John; McCrary, Justin, and Sanbonmatsu, Lisa, "Constructive Proposals for Dealing with Attrition: An Empirical Example," *NBER Working Paper*, July 2006.
- Eckel, Catherine, and Grossman, Philip J., "Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method on Dictator Experiments," *Experimental Economics*, 2, 2000, 107-120.
- Garvey, Paul R.; Book, Stephen A., and Covert, Raymond P. *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective* (Boca Raton, FL: CRC Press, Second Edition, 2015).
- Glöckner, Andreas, and Pachur, Thorsten, "Cognitive Models of Risky Choice: Parameter Stability and Predictive Accuracy of Prospect Theory," *Cognition*, 123, 2012, 21-32.
- Goldstein, Daniel G.; Johnson, Eric J., and Sharpe, William F., "Choosing Outcomes versus Choosing Products: Consumer-Focused Retirement Investment Advice," *Journal of Consumer Research*, 35, 2008, 440-456.
- Grasdal, Astrid, "The Performance of Sample Selection Estimators to Control for Attrition Bias," *Health Economics*, 10, 2001, 385-398.
- Harrison, Glenn W., "Randomisation and Its Discontents," *Journal of African Economies*, 20(4), 2011, 626-652.
- Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109(2), 2007, 341-368.
- Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," *Journal of Economic Behavior and Organization*, 70(3), June 2009, 498-507.
- Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).
- Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1009-1055.
- Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).
- Harrison, Glenn W., and Swarthout, J. Todd, "Cumulative Prospect Theory in the Lab: A Reconsideration," *Working Paper 2016-05*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.
- Hausman, Jerry A., and Wise, David A., "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, 47(2), March 1979, 455-473.

- Heckman, James J., "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 1976, 5, 475-492.
- Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), January 1979, 153-162.
- Heckman, James J., and Robb, Richard, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data* (New York: Cambridge University Press, 1985).
- Heckman, James J., and Smith, Jeffrey A., "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9(2), Spring 1995, 85-110.
- Heckman, James J.; Smith, Jeffrey A., and Taber, Christopher, "Accounting for Dropouts in Evaluations of Social Programs," *Review of Economics and Statistics*, 80(1), February 1998, 1-14.
- Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.
- Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 2002, 1644-1655.
- Horowitz, Joel L., and Manski, Charles F., "Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, March 2000, 77-84.
- Hotz, V. Joseph, "Designing an Evaluation of JTPA," in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs* (Cambridge: Harvard University Press, 1992).
- Kagel, John H.; Battalio, Raymond C., and Walker, James M., "Volunteer Artifacts in Experiments in Economics: Specification of the Problem and Some Initial Data from a Small-Scale Field Experiment," in V.L. Smith (ed.), *Research in Experimental Economics* (Greenwich, CT: JAI Press, 1979, volume 1).
- Keane, Michael P., "A note on identification in the multinomial probit model," *Journal of Business and Economic Statistics*, 10, 1992, 193-200.
- Kramer, Michael, and Shapiro, Stanley, "Scientific Challenges in the Application of Randomized Trials," *Journal of the American Medical Association*, 252, November 16, 1984, 2739-2745.
- Lee, David, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 2009, 76, 1071-1102.
- Lee, Lung-Fei, "Semiparametric maximum likelihood estimation of polychotomous and sequential choice models," *Journal of Econometrics*, 65, 1995, 381-428.

- Leon, Andrew C.; Mallinckrodt, Craig H.; Chuang-Stein, Christy; Archibald, Donald G.; Archer, Graeme E., and Chartier, Kevin, "Attrition in Randomized Controlled Clinical Trials: Methodological Issues in Psychopharmacology," *Biological Psychiatry*, 59, 2006, 1001-1005.
- Ludwig, Jens; Duncan, Greg J.; Gennetian, Lisa A.; Katz, Lawrence F.; Kessler, Ronald C.; Kling, Jeffrey R., and Sanbonmatsu, Lisa, "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity," *American Economic Review (Papers & Proceedings)*, 103(3), 2013, 226-231.
- Meng, Chun-Lo, and Schmidt, Peter, "On the cost of partial observability in the bivariate probit model," *International Economic Review*, 26, 1985, 71-85.
- Orr, Larry; Feins, Judith D. Feins; Jacob, Robin; Beecroft, Erik; Sanbonmatsu, Lisa; Katz, Lawrence F.; Liebman, Jeffrey B., and Kling, Jeffrey R., "Moving to Opportunity Interim Impacts Evaluation," *Final Report*, U.S. Department of Housing and Urban Development, 2003.
- Philipson, Tomas, "Data Markets and the Production of Surveys," *Review of Economic Studies*, 64, 1997, 47-72.
- Philipson, Tomas, "Data Markets, Missing Data, and Incentive Pay," *Econometrica*, 69, 2001, 1009-1111
- Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66, 1998, 497-527.
- Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.
- Sanbonmatsu, Lisa; Ludwig, Jens; Katz, Lawrence F.; Gennetian, Lisa A.; Duncan, Greg J.; Kessler, Ronald C.; Adam, Emma; McDade, Thomas W., and Lindau, Stacy Tessler, *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation* (Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, 2011; www.huduser.org/publications/pdf/MTOFHD_fullreport_v2.pdf).
- Slonim, Robert; Wang, Carmen; Garbino, Ellen, and Merrett, Danielle, "Opting-in: Participation Bias in Economic Experiments," *Journal of Economic Behavior & Organization*, 90, 2013, 43-70.
- Smidts, Ale, "The Relationship between Risk Attitude and Strength of Preference: A Test of Intrinsic Risk Attitude," *Management Science*, 43(3), 1997, 357-370.
- Train, Kenneth, *Discrete Choice Models with Simulation* (Cambridge, UK: Cambridge University Press, Second Edition, 2009).
- Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representations of Uncertainty," *Journal of Risk & Uncertainty*, 5, 1992, 297-323.
- Van de Ven, Wynand P.M.M., and Van Praag, Bernard M.S., "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection," *Journal of Econometrics*, 17, 1981, 229-252.

Vella, Francis, “Estimating Models with Sample Selection Bias: A Survey,” *Journal of Human Resources*, 33(1), 1998, 127-169.

Wilcox, Nathaniel T., “Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison,” in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data* (Cambridge, USA: MIT Press, Second Edition, 2010).

Zeisberger, Stefan; Vrecko, Dennis, and Langer, Thomas, “Measuring the Time Stability of Prospect Theory Preferences,” *Theory and Decision*, 72, 2012, 359-386.

Appendix A: Instructions (WORKING PAPER)

We document the instructions for the risk aversion task that were given in hard copy to the subjects and a typical screen shot of the decision task. The original Danish version of the manuscript is available on request. The instructions were in 14-point font, printed on A4 paper, and handed out in laminated form.

Task L

In this task you will make a number of choices between two options labeled “A” and “B”. An example of your task is shown on the right. You will make all decisions on a computer.

All decisions have the same format. In the example on the right Option A pays 60 kroner if the outcome of a roll of a ten-sided die is 1, and it pays 40 kroner if the outcome is 2-10. Option B pays 90 kroner if the outcome of the roll of the die is 1 and 10 kroner if the outcome is 2-10. All payments in this task are made today at the end of the experiment.

We will present you with 40 such decisions. The only difference between them is that the probabilities and amounts in Option A and B will differ.

You have a 1-in-10 chance of being paid for one of these decisions. The selection is made with a 10-sided die. If the roll of the die gives the number 1 you will be paid for one of the 40 decisions, but if the roll gives any other number you will not be paid. If you are paid for one of these 40 decisions, then we will further select one of these decisions by rolling a 4-sided and a 10-sided die. A third die roll with a 10-sided die determines the payment for your choice of Option A or B. When you make your choices you will not know which decision is selected for payment. You should therefore treat each decision as if it might actually count for payment.

If you are being paid for one of the decisions, we will pay you according to your choice in the selected decision. You will then receive the money at the end of the experiment.

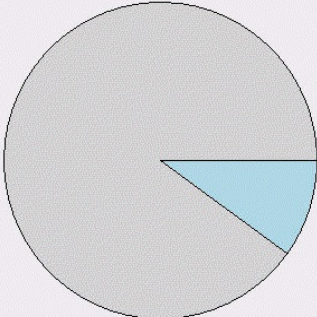
Before making your choices you will have a chance to practice so that you better understand the consequences of your choices. Please proceed on the computer to the practice task. You will be paid in caramels for this practice task.

Typical screen shot

ID: 1234

Decision number 1 out of 40

Option A



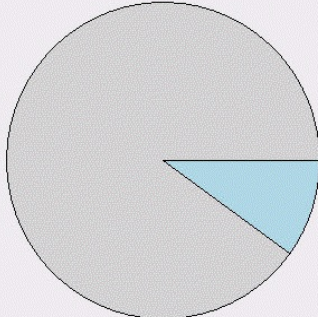
\$2000 if the number on the die is 1

\$1600 if the number on the die is 2 to 10

Select A

Continue

Option B



\$3850 if the number on the die is 1

\$100 if the number on the die is 2 to 10

Select B

Appendix B: Additional Estimations (WORKING PAPER)

**Table B1: Estimates of EUT Parameters
with No Controls for Sample Selection and Attrition**

(Log-simulated likelihood = 9556 for 25,555 observations on 413 subjects in wave 1 and 182 subjects in wave 2 using 100 Halton draws.)

Variable	Estimate	Standard Error	<i>p</i> -value	95% Confidence Interval	
<i>Mean of r parameter in wave 1</i>					
RAfirst	-0.014	0.066	0.829	-0.143	0.115
RAhigh	0.066	0.025	0.009	0.017	0.115
constant	0.537	0.058	0.000	0.424	0.650
<i>Mean of r parameter in wave 2</i>					
RAfirst	-0.101	0.128	0.428	-0.352	0.149
RAhigh	0.047	0.035	0.185	-0.022	0.116
constant	0.660	0.152	0.000	0.362	0.957
<i>Standard deviations and correlation coefficient of r parameters in wave 1 and wave 2</i>					
σ_{r1}	0.698	0.056	0.000	0.587	0.808
σ_{r2}	0.549	0.049	0.000	0.453	0.646
ρ_{r1r2}	0.575	0.079	0.000	0.419	0.730
<i>Test for stability of predicted group means for r parameter</i>					
Δ Base	0.128	0.126	0.331	-0.125	0.370
Δ RAhigh	0.104	0.123	0.396	-0.136	0.345
Δ RAfirst	0.036	0.058	0.542	-0.079	0.150

Notes: Group means are predicted using the estimated mean function for r parameter. Δ Base tests whether the between-wave difference in constant is significant. Δ RAhigh (Δ RAfirst) tests whether the between-wave difference in constant + RAhigh (constant + RAfirst) is significant.

**Table B2: Estimates of the RDU Parameters
with No Controls for Sample Selection and Attrition**

(Log-simulated likelihood = 8487 for 25,555 observations on 413 subjects in wave 1 and 182 subjects in wave 2 using 100 Halton draws.)

Variable	Estimate	Standard Error	<i>p</i> -value	95% Confidence Interval	
<i>Mean of r parameter in wave 1</i>					
RAfirst	0.140	0.163	0.391	-0.180	0.459
RAhigh	0.053	0.035	0.135	-0.017	0.122
constant	0.484	0.152	0.001	0.186	0.783
<i>Mean of r parameter in wave 2</i>					
RAfirst	-0.006	0.116	0.957	-0.233	0.221
RAhigh	0.028	0.045	0.536	-0.060	0.116
constant	0.656	0.094	0.000	0.472	0.840
<i>Standard deviations and correlation coefficient of r parameters in wave 1 and wave 2</i>					
σ_{r1}	0.642	0.067	0.000	0.510	0.775
σ_{r2}	0.553	0.058	0.000	0.439	0.667
ρ_{r1r2}	0.546	0.045	0.000	0.458	0.635
<i>Mean of φ parameter in wave 1</i>					
RAfirst	0.408	0.360	0.257	-0.298	1.114
RAhigh	0.083	0.104	0.424	-0.121	0.288
constant	1.639	0.268	0.000	1.114	2.165
<i>Mean of φ parameter in wave 2</i>					
RAfirst	0.614	0.654	0.348	-0.668	1.896
RAhigh	0.029	0.174	0.865	-0.312	0.371
constant	2.438	0.477	0.000	1.502	3.374
<i>Standard deviations and correlation coefficient of φ parameters in wave 1 and wave 2</i>					
$\sigma_{\varphi1}$	3.338	0.578	0.000	2.205	4.471
$\sigma_{\varphi2}$	8.254	2.983	0.006	2.409	14.100
$\rho_{\varphi1\varphi2}$	0.591	0.050	0.000	0.493	0.689

Other correlation coefficients

$\varrho_{r1\varphi1}$	0.271	0.034	0.000	0.205	0.338
$\varrho_{r1\varphi2}$	0.128	0.030	0.000	0.070	0.187
$\varrho_{r2\varphi1}$	0.277	0.038	0.000	0.203	0.351
$\varrho_{r2\varphi2}$	0.224	0.035	0.000	0.155	0.292

Test for stability of predicted group means for r parameter

ΔBase	0.172	0.107	0.109	-0.038	0.381
ΔRAhigh	0.146	0.093	0.115	-0.035	0.328
$\Delta\text{RAfirst}$	0.026	0.075	0.731	-0.121	0.172

Test for stability of predicted group means for φ parameter

ΔBase	0.799	0.442	0.071	-0.067	1.664
ΔRAhigh	0.745	0.448	0.097	-0.134	1.624
$\Delta\text{RAfirst}$	1.004	0.559	0.072	-0.092	2.100

Notes: Group means are predicted using the estimated mean function for each parameter. ΔBase tests whether the between-wave difference in constant is significant. ΔRAhigh ($\Delta\text{RAfirst}$) tests whether the between-wave difference in constant + RAhigh (constant + RAfirst) is significant.