ESTIMATING RISK ATTITUDES IN DENMARK:

A FIELD EXPERIMENT

by

GLENN W. HARRISON, MORTEN I. LAU AND E. ELISABET RUTSTRÖM *

March 2004

Abstract. We estimate individual risk attitudes using controlled experiments in the field in Denmark. These risk preferences are elicited by means of field experiments involving real monetary rewards. The experiments were carried out across Denmark using a representative sample of 253 people between 19 and 75 years of age. Risk attitudes are estimated for various individuals differentiated by socio-demographic characteristics such as income and age. Our results indicate that the average Dane is risk averse, and that risk neutrality is an inappropriate assumption to apply. We also find that risk attitudes do vary significantly with respect to several important socio-demographic variables. These conclusions are robust to the use of relatively flexible specifications of risk preferences. When individual characteristics of the sample are ignored, relative risk aversion appears not to be constant over the domain of income considered here, and rises rapidly as income increases above "small" amounts. However, relative risk aversion appears to be constant when one corrects for individual heterogeneity, although there is considerable uncertainty in the characterization of risk attitudes for low stakes.

Keywords: Risk preferences, field experiments, heterogeneity.

JEL codes: D00

* Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström), and Centre for Economic and Business Research, Copenhagen, Denmark (Lau). E-mail contacts: GHARRISON@BUS.UCF.EDU, MOL@CEBR.DK, and ERUTSTROM@BUS.UCF.EDU. Rutström thanks the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/MRI 9871019 and NSF/POWRE 9973669, and Harrison and Lau thank the Danish Social Science Research Council for research support under project #24-02-0124. Steffen Andersen provided superb research assistance and comments throughout. Supporting data and instructions are stored at the ExLab Digital Archive located at http://exlab.bus.ucf.edu.

Assumptions about risk attitudes play a central role in the analysis of major economic decisions such as education, employment, health care, and retirement. Whenever costs and benefits for a household or individual are uncertain, it is essential that one calculate certainty equivalents in order to undertake meaningful comparisons. In most cases welfare analysts implicitly assume risk neutrality as the basis for these calculations, although it is *a priori* plausible that risk aversion might be very high for some vulnerable segments of the population. In fact, since risk attitudes are a reflection of subjective preferences, one would expect *a priori* that they would differ across individuals.[1]

We elicit measures of individual risk attitudes from subjects in Denmark in order to test three substantive hypotheses. The first hypothesis is that *risk attitudes differ significantly from risk neutrality*, such that the implicit assumption in cost-benefit analysis should be reviewed.[2] The second hypothesis is that there are *identifiable segments of the population that exhibit significant differences in risk attitudes*, such that analysts should allow for heterogeneity when they can identify those segments. The third hypothesis is that *relative risk aversion is not constant* with respect to the income levels of the lottery prizes considered, such that one should avoid popular constant relative risk aversion specifications for policies defined over non-trivial income changes.

We use survey questions with real monetary rewards to elicit risk attitudes and demonstrate the methodological complementarity between lab and field experiments.[3] The survey questions are based on those designed by Holt and Laury [2002], who elicited risk attitudes for university students using controlled laboratory experiments. We apply extended versions of their experimental

---

[1] We elicit risk attitudes for individuals. To the extent that the characteristics of individuals are used to define "representative households," we can refer to the individual and the household interchangeably. However, we remain agnostic concerning the way in which the individual risk attitudes of individual household members are aggregated into one household risk attitude.

[2] Following Rabin [2000], there are some specifications of expected utility theory for which a finding of risk aversion at these levels of income is incoherent. This argument does not apply if expected utility theory is defined over income earned during the experiment, rather than over terminal lifetime wealth. Appendix A reviews this argument, and its relevance for experimental studies of risk aversion.

[3] Following the taxonomy developed by Harrison and List [2003], we conduct artefactual field experiments.

procedures, but employ subjects that are normally encountered in field surveys. Our field experiments were carried out across Denmark for the Danish government, using a nationally representative sample of 253 people between 19 and 75 years of age.

Our results indicate that the average Dane is risk averse, and that risk neutrality is an inappropriate assumption to apply. We also find that risk attitudes do vary significantly with respect to several important socio-demographic variables. These general conclusions are robust to the use of relatively flexible specifications of risk preferences. Relative risk aversion appear to vary over the domain of income considered here, rising rapidly as income increases above "small" amounts, if one fails to account for individual heterogeneity in risk attitudes. We do account for that heterogeneity, and find that the assumption of constant relative risk aversion is acceptable over the domain of income considered. However, there is considerable uncertainty in the estimation of risk attitudes for small amounts of income. In that domain one cannot reject the hypothesis that subjects are risk neutral or even risk loving, simply because the precision of the estimates is so poor. Our results consistently support the need to recognize the heterogeneity of risk attitudes across individual subjects. This result has important implications for the characterization of risk attitudes in policy applications, theoretical modeling, and experimental economics.

In section 1 we review the logic of our experimental design. We propose several extensions of the basic laboratory procedure designed to elicit more precise responses and check for robustness to framing effects. These extensions provide several methodological improvements in the risk elicitation procedure, which are of independent interest. Section 2 explains the field experiments conducted, with additional details on procedures provided in Harrison, Lau, Rutström and Sullivan [2004]. Section 3 examines the results and relates them to those found in the existing literature. We also demonstrate how our design allows one to "calibrate" for possible framing effects, and evaluate relatively flexible functional forms for risk preferences.

At a methodological level, we demonstrate that it is possible to elicit risk attitudes in a field experiment that reflects the population of a country. We concede that Denmark is a remarkable country in which to recruit subjects and undertake field experiments – over 94% of the field subjects we recruited actually turned up for their sessions! The potential importance of eliciting risk attitudes for policy evaluation justifies the development of procedures to rigorously elicit risk attitudes as one component of large-scale surveys that are routinely conducted in many countries. Our procedures should serve as a "best-case" guide to such efforts in the future.

1. Experimental Design

A. The Basic Elicitation Procedure

Holt and Laury [2002] (HL) devise a simple experimental measure for risk aversion using a multiple price list (MPL) design.[4] Each subject is presented with a choice between two lotteries, which we call A or B. Table 1 illustrates the basic payoff matrix presented to subjects. The first row shows that lottery A offered a 10% chance of receiving $2 and a 90% chance of receiving $1.60. The expected value of this lottery, EV^A, is shown in the third-last column as $1.64, although the EV columns were not presented to subjects. Similarly, lottery B in the first row has chances of payoffs of $3.85 and $0.10, for an expected value of $0.48. Thus the two lotteries have a relatively large difference in expected values, in this case $1.17. As one proceeds down the matrix, the expected value of both lotteries increases, but the expected value of lottery B becomes greater than the expected value of lottery A.

The subject chooses A or B in each row, and one row is later selected at random for payout for that subject. The logic behind this test for risk aversion is that only risk-loving subjects would take lottery B in the first row, and only risk-averse subjects would take lottery A in the second last row. Assuming local non-satiation, the last row is simply a test that the subject understood the

[4] The MPL appears to have been first used in pricing experiments by Kahneman, Knetsch and Thaler [1990], and has been adopted in recent discount rate experiments by Coller and Williams [1999]. It has a longer history in the elicitation of hypothetical valuation responses in "contingent valuation" survey settings, as discussed by Mitchell and Carson [1989; p. 100, fn. 14]. The test devised by HL is closely related to one developed by Murnighan, Roth and Schoumaker [1988] to measure the degree of risk aversion of subjects in bargaining experiments.

instructions, and has no relevance for risk aversion at all. A risk neutral subject should switch from choosing A to B when the EV of each is about the same, so a risk-neutral subject would choose A for the first four rows and B thereafter.

These data may be analyzed using a variety of statistical models. Each subject made 10 responses. The responses can be reduced to a scalar if one looks at the *lowest* row in Table 1 at which the subject "switched" over from lottery A to lottery B.[5] This reduces the response to a scalar for each subject and task, but a scalar that takes on integer values between 0 and 10. Alternatively, one could study the effects of experimental conditions in terms of the constant relative risk aversion (CRRA) characterization, employing an interval regression model. The CRRA utility is defined as $U(y) = (y^{1-r})/(1-r)$, where r is the CRRA coefficient.[6] The dependant variable in the interval regression model is the CRRA interval that subjects implicitly choose when they switch from lottery A to lottery B. For each row of panel A in Table 1, one can calculate the implied bounds on the CRRA coefficient, and these are in fact reported by HL [2002; Table 3]. These intervals are shown in the final column of Table 1. Thus, for example, a subject that made 5 safe choices and then switched to the risky alternatives would have revealed a CRRA interval between 0.14 and 0.41, and a subject that made 7 safe choices would have revealed a CRRA interval between 0.68 and 0.97, and so on.

HL also utilize a variant of the Expo-Power (EP) utility function proposed by Saha [1993], which is more general than the CRRA characterization. The EP function is defined as $u(y) = [1-\exp(-\alpha y^{1-r})]/\alpha$, where y is income and α and r are parameters to be estimated. Relative risk aversion (RRA) is then $r + \alpha(1-r)y^{1-r}$. So RRA varies with income if $\alpha \neq 0$. This function nests CRRA (as α tends to 0) and CARA (as r tends to 0). HL estimate this function assuming that every subject has the same risk preference. They rely on a "noise parameter" to accommodate the obvious differences

[5] Some subjects switched several times, but the minimum switch point is always well-defined. It turns out not to make much difference how one handles these "multiple switch" subjects, but our analysis and the analysis of HL consider the effect of accounting for them in different ways explained below.

[6] With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(m) = \ln(m)$.

in risk choices across subjects, but do not allow risk preferences to vary with observable socio-demographic characteristics as we do later.

Our approach is to use the CRRA specification for basic analysis of results, and then examine robustness using EP specifications.

B. Extensions

We expand the HL design, with some simple modifications to allow a richer characterization of the utility function and the reliability of the elicitation procedure.

Variations in the Income Domain

The design used by HL called for each subject to be given choices over four lottery prizes and for there to be one major scale change for all real payoffs. We also want to allow for changes in the value of prizes, so that we have data for the same subject over more than four prizes and can generate better characterizations of their risk attitudes.

We undertake four separate risk aversion tasks with each subject, each with different prizes designed so that all 16 prizes span the range of income that we seek to estimate risk aversion over. Ideally, we would have a roughly even span of prizes so that we can evaluate the utility function for the individual at different income levels and know that there were some response at or near that level. The four sets of prizes are as follows, in Danish kroner (DKK), with the two prizes for lottery A listed first and the two prizes for lottery B listed next: (A1: 2000, 1600; B1: 3850, 100), (A2: 2250, 1500; B2: 4000, 500), (A3: 2000, 1750; B3: 4000, 150), and (A4: 2500, 1000; B4: 4500, 50). At the time of the first phase of the experiments, the exchange rate was approximately 6.55 DKK per U.S. dollar, so these prizes range from approximately $7.65 to $687.

This set of prizes generates an array of possible CRRA values. For example, set 1 generates CRRA intervals at the switch points of -1.71, -0.95, -0.49, -0.14, 0.15, 0.41, 0.68, 0.97 and 1.37. The other sets generate different CRRA intervals, such that all four sets span 36 distinct CRRA values between -1.84 and 2.21, with roughly 60% of the CRRA values reflecting risk aversion.[7] Any scaling of the prizes that is common within a set will preserve the implied CRRA coefficients, so this design could also be used with smaller or larger payoffs.

We ask the subject to respond to all four risk aversion tasks in the same order and then randomly decide which one to play out. Budget constraints precluded paying all subjects, so each subject is given a 10% chance of actually receiving the payment associated with his decision.

Iterating the MPL

It is possible to extend the MPL to allow more refined elicitation of the true risk attitude, and yet retain the transparency of the incentives of the basic MPL. We do so in the form of a computerized variant on the basic MPL format which we call an Iterative MPL (iMPL).

The basic MPL is the standard format in which the subject sees a fixed array of paired options and chooses one for each row. It allows subjects to switch back and forth as they like, and has already been used in many experiments. The iMPL format extends this by first asking the subject to simply choose the row at which he wants to first switch from option A to option B, assuming monotonicity of the underlying preferences to automatically fill out the remaining choices. The second extension of the MPL format is to then allow the individual to make choices from refined options within the option last chosen. That is, if someone decides at some stage to switch from option A to option B between probability values of 0.1 and 0.2, the next stage of an iMPL would then prompt the subject to make more choices *within* this interval, to refine the values elicited.[8] Figures 1 and 2 illustrate Level 1 and Level 2, respectively, of an iMPL. In Level 1 the illustrative subject first chooses B when the interest rate is between 0.3 and 0.4, so that Level 2

[7] The second set generates CRRA values of -1.45, -0.72, -0.25, 0.13, 0.47, 0.80, 1.16, 1.59 and 2.21; the third set generates values of -1.84, -1.101, -0.52, -0.14, 0.17, 0.46, 0.75, 1.07 and 1.51; and the fourth set generates values of -0.75, -0.32, -0.05, 0.16, 0.34, 0.52, 0.70, 0.91 and 1.20.

[8] If the subject always chooses A, or indicates Indifference for any of the decision rows, there are no additional decisions required and the task is completed.

presents the subject with 11 more choices within the interval 0.3 to 0.4.[9]

The iMPL uses the same incentive logic as the MPL. After making all responses, the subject has one row from the first table selected at random by the experimenter. In the MPL that is all there is. In the iMPL, that is all there is if the row selected at random by the experimenter is *not* the one that the subject switched at in Level 1. If it *is* the row that the subject switched at, another random draw is made to pick a row in the Level 2 table, and so on.

As the subject iterates in the iMPL, the choices become more and more alike, by design. Hence one would expect that greater cognitive effort would be needed to discriminate between them. At some point we expect the subject to express indifference, which we account for in our analysis by only considering the interval over which the subject could (strictly) discriminate.

Framing Effects

A natural concern with the MPL and iMPL is that it might encourage subjects to pick a response in the middle of the table, independent of true valuations. There could be a psychological bias towards the middle, although that is far from obvious *a priori*.

One solution to this concern which we find unattractive is to randomize the order of the rows. This is popular in some experimental studies in psychology and economics which elicit discount rates and risk attitudes using the MPL.[10] We find it unattractive for two reasons. First, if there is a purely psychological anchoring effect towards the middle, this will do nothing but add noise to the responses. Second, the valuation task is fundamentally harder from a cognitive perspective if one shuffles the order of valuations across rows. This harder task may be worthy of study, but is a needless confound for our inferential purposes.[11]

---

[9] The iterative format has some "smarts" built into it: when the values being elicited drop to some specified perceptive threshold (e.g., set to a 1-in-100 die throw), the iMPL collapses down to an endogenous number of final rows and the elicitation task stops iterating after those responses are entered.

[10] See Kirby and Maraković [1996], Kirby, Petry and Bickel [1999] and Eckel, Johnson and Montmarquette [2002].

[11] A parallel issue arises in the use of incentive-compatible auction institutions to elicit valuations for private goods. One might be interested in testing if subjects behave as if they understand that truthfully

Framing effects can be relatively easily tested for by varying the cardinal scale of the basic MPL table, or by varying the number of intervals within a given cardinal range. If there is an effect on responses, it will be easy to identify statistically and then to correct for in the data analysis. We would not be surprised to find framing effects of this kind. They do not necessarily indicate a failure of the traditional economic model, so much as a need to recognize that subjects in a lab setting use all available information to identify a good valuation for a commodity (Harrison, Harstad and Rutström [2004]). Thus it is important to be able to estimate the quantitative effect of certain frames and then allow for them in subsequent statistical analysis.

We devise a test for framing effects by varying the cardinal scale of the MPL used in the risk aversion task. Two asymmetric frames are developed: the *skewHI* treatment offers initial probabilities of (0.3, 0.5, 0.7, 0.8, 0.9 and 1), while *skewLO* offers initial probabilities of (0.1, 0.2, 0.3, 0.5, 0.7, and 1). This treatment yields 6 decision rows in Level 1 of the iMPL, as opposed to the 10 rows in the symmetric frame.[12] As suggested by the treatment names, *skewLO* (*skewHI*) is intended to skew responses to be lower (higher) probabilities if subjects pick in the middle.

2. The Danish Experiments

A. Sampling Procedures

The sample for the field experiments was designed to generate a representative sample of the adult Danish population. There were six steps in the construction of the sample,[13] essentially following those employed in HLW:

---

revealing their valuation is in their best interest, or one might be interested in using that feature of the institution to encourage truthful revelation. In the latter case it would be appropriate to tell subjects the (correct) logic underlying the incentive-compatibility of the institution, but that would obviously be an unfortunate design choice in the former case.

[12] The skewed frames does interact with the implementation of the iMPL. In the symmetric frame, all intervals are 10 probability points wide, so that a second level is all that is needed to bring subject choices down to precise intervals of 1 probability point. In the skewed frames, however, because the intervals vary in size, a third level is required to bring choices down to this level of precision, and the number of decision rows in Level 3 depends on the width of the interval in Level 1 at which the subject switches.

[13] Further details are provided in Harrison, Lau, Rutström and Sullivan [2004].

- First, a random sample of 25,000 Danes was drawn from the Danish Civil Registration Office in January 2003. Only Danes born between 1927 and 1983 were included, thereby restricting the age range of the target population to between 19 and 75. For each person in this random sample we had access to their name, address, county, municipality, birth date, and sex. Due to the absence of names and/or addresses, 28 of these records were discarded.
- Second, we discarded 17 municipalities (including one county) from the population, due to them being located in extraordinarily remote locations. The population represented in these locations amounts to less than 2% of the Danish population, or 493 individuals in our sample from the civil registry.
- Third, we assigned each county either 1 session or 2 sessions, in rough proportionality to the population of the county. In total we assigned 20 sessions. Each session consisted of two sub-sessions at the same locale and date, one at 5pm and another at 8pm, and subjects were allowed to choose which sub-session suited them best.
- Fourth, we divided 6 counties into two sub-groups because the distance between some municipalities in the county and the location of the session would be too large. A random draw was made between the two sub-groups and the location selected, where the weights reflect the relative size of the population in September 2002.
- Fifth, we picked the first 30 or 60 randomly sorted records within each county, depending on the number of sessions allocated to that county. This provided a sub-sample of 600.
- Sixth, we mailed invitations to attend a session to the sub-sample of 600, offering each person a choice of times for the session.[14] Response rates were low in some counties, so

[14] The initial letter of invitation included an answer form and a prepaid envelope, and the subject was asked to answer within one week. The same day we received the answer form, a reply letter was sent confirming their participation in the meeting at the given location, date and time. Every recruited subject was reminded by mail or phone within a week of the meeting. Both procedures were used for the first three sessions, and attendance was almost 100% at these sessions. We reminded subjects by mail for the remaining sessions because this procedure is more convenient.

another 64 invitations were mailed out in these counties to newly drawn subjects.[15] Everyone that gave a positive response was assigned to a session, and our recruited sample was 268.[16]

Attendance at the experimental sessions was extraordinarily high, including 4 persons who did not respond to the letter of invitation but showed up unexpectedly and participated in the experiment. Four persons turned up for their session, but were not able to participate in the experiments.[17]

These experiments were conducted between June 2 and June 24, 2003, and a total of 253 subjects participated in the experiments.[18] Sample weights for the subjects in the experiment can be constructed using the sample design, as explained by Harrison, Lau, Rutström and Sullivan [2004].

B. Conduct of the Sessions

To minimize travel times for subjects, we reserved hotel meeting rooms in convenient locations across Denmark in which to conduct sessions.[19] Because the sessions lasted for two hours, light refreshments were provided. Participants met in groups of no more than 10. To conduct

[15] An additional 45 and 19 invitations were sent out in the second and third wave, respectively. The first wave of invitations were sent out four weeks before the first session was scheduled, and we asked people to reply within one week. The second and third waves of invitations were sent out two and three weeks after the first wave, respectively.

[16] The response rate was 42.5% for the first wave, 20.0% percent for the second wave, and 22.1% for the third wave.

[17] The first person suffered from dementia and could not remember the instructions; the second person was a 76 year old woman who was not able to control the mouse and eventually gave up; the third person had just won a world championship in sailing and was too busy with interviews to stay for two hours; and the fourth person was sent home because too many people showed up (one person came unexpected, and we had only ten laptops available at that session).

[18] Certain events might have plausibly triggered some of the no-shows: for example, 3 men did not turn up on June 11, 2003, but that was the night that the Danish national soccer team played a qualifying game for the European championships against Luxembourg that was not scheduled when we picked session dates.

[19] It is possible to undertake experiments over the web with a large sample of subjects drawn from the population. Kapteyn and Teppa [2003] illustrate how one can elicit hypothetical responses to elicit time preferences using a panel of 2,000 Dutch households connected by home computer to surveys. Although not concerned with risk and time preferences directly, Hey [2002] illustrates how one can augment such electronic panel surveys with real experiments. Donkers and van Soest [1999] elicit hypothetical risk and time preferences from pre-existing panels of Dutch households being surveyed for other reasons.

computerized experiments in the field, it was cost-effective to purchase laptop computers and transport them to the meeting sites. It was not necessary to network the computers for these experiments; the program ran independently on each computer and results for each subject saved on that laptop. Each subject was identified by a unique ID number. For the randomization procedures, two bingo cages were used in each session, one containing 100 balls, and the other containing 3 to 11 balls, depending on the number of decision rows in the iMPL used in different treatments. We found two bingo cages to be the most transparent and convenient way to generate random outcomes in the experiments.

To begin the sessions, subjects were welcomed and reminded that they were to be paid 500 DKK for their participation to cover travel costs as long as they were able to stay for the full two hours required for the experiment. Anyone who was not able to stay for the full two hours was paid 100 DKK and excused from the experiment. The experimenters then asked for a volunteer to inspect and verify the bingo cages and number of bingo balls.

Instructions for the experiment were provided on the computer screens, and subjects read through the instructions while the experimenter read them aloud. The experimenters followed the same script and procedures for each session, documented in Harrison, Lau, Rutström and Sullivan [2004].

The experiment was conducted in four parts. Part I consisted of a questionnaire collecting subjects' socio-demographic characteristics. Specifically, we collected information on age, gender, size of town the subject resided in, type of residence, primary occupation during the last 12 months, highest level of education, household type (viz., marital status and presence of younger or older children), number of people employed in the household, total household income before taxes, disposable household income, whether the subject is a smoker, and the number of cigarettes smoked per day. Part IV consisted of another questionnaire which elicits information on the subject's financial market instruments, and probes the subject for information on their expectations about their future economic conditions and their own future financial position. The questionnaires are rather long, so we chose to divide them across Parts I and IV in order to reduce subject fatigue and boredom. Part II consisted of the four risk aversion tasks, and Part III presented subjects with the six discount rate tasks similar to those developed in Harrison, Lau and Williams [2002].

The four risk aversion tasks incorporate the incentive structure and assigned frames as described earlier. After subjects completed the four tasks, several random outcomes were generated in order to determine subject payments. For all subjects, one of the four tasks was chosen, then one of the decision rows in that task was chosen. For those subjects whose decision at that row led to Level 2, another random draw was required to choose a decision row in Level 2, and yet another random draw was required should that decision have led a subject to Level 3 in the iMPL. To maintain anonymity we performed the draws without announcing to which subjects it would apply. In the case where a subject indicated Indifference for the chosen decision row, another random draw determined whether the subject received the results from Lottery A or Lottery B. At this point all subjects knew whether they were playing Lottery A or Lottery B, and another random draw determined whether subjects were to receive the high payment or the low payment. Finally, a 10-sided die was rolled for each subject. Any subject who received a roll of "0" received actual payment according to that final outcome. All payments were made at the end of the experiment.

A significant amount of time was spent training subjects on the iMPL and the randomization procedures in Part II of the experiment. Subjects were given handouts containing examples of Level 1 and Level 2 of an iMPL that had been filled in. The training exercise explained the logic of the iMPL and verified that subjects were able to correctly fill in an iMPL as shown in the handout. Next, the experimenters illustrated the random procedures necessary to reach a final lottery outcome for each possible choice in the chosen Level 1 decision row. Finally, a single trainer task was conducted in which payments were in the form of candies. The ten-sided die was rolled for each subject, and candies were given to each subject who received a roll of "0."

3. Results

We present results by answering four questions. First, what is the general level of risk aversion in the Danish population, at least over the domain of income considered here? Specifically, is risk neutrality an acceptable hypothesis? Second, do risk attitudes vary with observable demographics? Related to the possible effect of demographics, we can also ask if there are significant effects on elicited risk attitudes from the task frame or task order. Third, how plausible is it to assume that relative risk aversion is constant, again over the domain of income considered here? Fourth, how does the characterization of risk change if one considers variants from the standard expected utility theory model, such as "stochastic choice behavior" or non-standard "probability weighting functions"?

A. Risk Aversion

Figure 3 shows the observed distribution of risk attitudes in our sample, using the raw mid-point of the elicited interval in the *final* iteration stage of the iMPL. This distribution reflects the symmetric menu treatment, which is the appropriate baseline to evaluate the asymmetric menu treatments. For this specification of CRRA, a value of 0 denotes risk neutrality, negative values indicate risk-loving, and positive values indicate risk aversion. Thus we see clear evidence of risk aversion: the mean CRRA coefficient is 0.64. This distribution is consistent with comparable estimates obtained in the United States, using college students and an MPL design, by Holt and Laury [2002] and Harrison, Johnson, McInnes and Rutström [2003a][2003b].[20]

Very few subjects exhibit any risk-preference. Friedman [1981] argues that subjects should never exhibit risk-loving behavior in an artefactual experiment even if they are risk lovers, since they have cheaper ways to purchase uncertainty in naturally occurring markets (e.g., purchase of lottery tickets, or trips to a casino). Our findings are consistent with this "censoring" hypothesis, although

[20] It is also consistent with estimates derived from models that entail joint tests of additional hypotheses, such as "noisy" equilibrium behavior in first-price auctions or normal form games (Goeree, Holt and Palfrey [2002][2003])

it could just be that nobody in our sample was a risk lover before taking this censoring possibility into account.

The unconditional data indicates that there is an effect on elicited risk aversion from the framing treatments. Figure 4 displays these data in a manner that allows one to easily compare the effects of the treatment. The *skewLO* treatment resulted in an average CRRA of 0.43, and the *skewHI* treatment resulted in an average CRRA of 0.95, each in the direction predicted *a priori*. Both are consistent with the conclusion that subjects are risk averse, and the rejection of the hypothesis of risk neutrality. Further comparison of the effect of these treatments on elicited risk aversion requires that we condition on the observed differences in the samples assigned to each treatment. Although subjects were randomly assigned to treatment, our samples are not large enough to be able to draw reliable conclusions solely on the basis of randomization (nor were they designed to).

The precision with which one estimates risk attitudes is of great significance for the interpretation of experimental data. If experimenters are unable to pin down the risk attitudes of subjects, then tests that rely on risk attitudes have correspondingly lower statistical power. To take one important example, consider the common ratio lottery pair widely used to evaluate EUT: the parameters employed by Cubitt, Starmer and Sugden [1998] result in indifference between two of the lotteries when subjects have a CRRA of 0.45. If the analyst cannot rule out the possibility that the subject has a CRRA of 0.45, no amount of experimental data can disprove the hypothesis that subjects are indifferent between the two lotteries.[21] For another example, consider the lottery pairs used in the famous preference reversal experiments of Grether and Plott [1979]. In this case, by design, the subject is indifferent between the lotteries if risk neutral. Hence, if one cannot rule out risk neutrality, one cannot rule out the hypothesis that EUT has no predictions over these choice tasks.[22]

[21] Harrison, Johnson, McInnes and Rutström [2003b] pursue this example in more detail, eliciting risk attitudes from the same subjects that they elicit lottery choices from. Their *in-sample* design allows relatively precise statements about the validity of EUT.

[22] We do not make the assumption that indifference would imply a specific 50/50 split in observed choices, since it does not follow from any standard axiom of EUT. Nor does it make sense: all sorts of non-

B. Heterogeneity

In order to assess the importance of demographics on risk attitudes, we applied regression models that condition on observable characteristics of the subjects. Table 2 provides the definitions of the explanatory variables and summary statistics. Table 3 displays the results from estimating a panel interval regression model of the elicited CRRA values. This model uses panel data since each subject provided four responses, one for each stake condition.[23] Unobserved individual effects are modeled using a random-effects specification.[24]

We first consider the marginal effect of individual demographics, holding constant the average value of other demographics, and then consider the joint effect of demographics.

<u>Marginal Effects</u>

We find no effect of sex on risk aversion. The absence of an effect of sex is noteworthy, since it has been intensively studied using related experimental and survey methods, and has even been the focus of theorizing about the role of evolution in forming preferences.[25]

---

economic biases might be expected to have a particularly strong role determining choice patterns when economic factors play no role.

[23] Several checks are undertaken for the specification. First, collapsing the intervals down to their mid-point allows a comparison of random-effects and fixed-effects specifications, and a Hausman test that the random-effects specification is consistent. There is no evidence that the random-effects specification is inconsistent, using the 90% or so of responses that had an interval less than 7 percentage points. Second, a Breusch-Pagan test of the null hypothesis that there is no variance in the unobserved individual random effects is convincingly rejected. Third, since potentially fragile numerical quadrature methods are used to estimate this specification, we checked for numerical stability as the number of quadrature points is varied, and there was no evidence of instability in the log-likelihood or any of the individual coefficients. These specification tests are performed for all of our panel models with similar results.

[24] The term "unobserved individual effect" refers to the fact that we do observe that certain observations come from the same individual, but that they might reflect characteristics of the individual that we have not observed.

[25] Levin, Snyder and Chapman [1988] and Powell and Ansic [1997] illustrate the experimental studies undertaken in a settings in which the task was not abstract but there were no real earnings by subjects. Harbaugh, Krause and Vesterlund [2002] and Holt and Laury [2002] conduct abstract experiments with real rewards, and find no significant sex effects on elicited risk aversion when stakes are non-trivial. Schubert, Brown, Gysler and Brachinger [1999] conduct abstract and non-abstract experiments with real rewards, and conclude that women do appear to be more risk averse than men in abstract tasks in the gain frame, but that this effect disappears with context. Unfortunately, they employed the Becker-DeGroot-Marschak procedure for eliciting certainty-equivalents, which is known to have poor incentive properties for experimental subjects.

We do find an effect from age on risk attitudes. Younger individuals, under the age of 30, tend to be more risk averse than those aged between 30 and 39, although the effect is not statistically significant. After 40 subjects become significantly less risk averse than those aged betwee 30 and 39. Again, we are controlling for other demographics that are plausibly associated with age, so these effects occur even after allowing for those factors.

Skilled workers with some post-secondary education have significantly higher aversion to risk. However, this is mitigated if the subject has experienced substantial higher education.

The variables *skewLO* and *skewHI* control for the frame used, as noted earlier. The *skewHI* treatment is statistically significant, consistent with the display in the bottom panel of Figure 4. Despite the unconditional estimates of mean CRRA differing from 0.64 to 0.43 when comparing the symmetric and *skewLO* treatments, after controlling for differences in the samples the estimates from Table 3 indicate that there is no significant effect from the *skewLO* treatment.[26]

We observe that there is an effect on the CRRA coefficient from varying the lottery prizes across the four tasks. There is a significant difference between Task 1 (the reference task for this statistical analysis) and the other three tasks. In particular, Task 2 is associated with higher CRRA responses, with a significant coefficient value of 0.32. The effect is smaller for Tasks 3 and 4 relative to Task 1, but still positive and significant. We therefore confirm the findings reported in Holt and Laury [2002] and Harrison, Johnson, McInnes, and Rutström [2003a][2003b] that the *relative* risk aversion coefficient is not constant in the stakes, although here we varied the stakes in a non-monotonic manner. One can either allow for this task effect with a CRRA characterization that conditions on it, as we do here, or explore more flexible specifications than CRRA that might

---

Jianakoplos and Bernasek [1998] examine data from the *U.S. Survey of Consumer Finances*, and conclude that single women are more risk averse in their financial choices than single men. Eckel and Grossman [2003] review these studies and several unpublished studies. Rubin and Paul [1979] and Robson [1996] offer evolutionary models of possible sex differences in risk aversion.

[26] Harrison, Lau, Rutström and Sullivan [2004] show that there is a significant effect on elicited risk attitudes from *both* skewness treatments on the *initial* stage of the iMPL, but that the iterations of the iMPL make that effect disappear for the skewLO treatment. Our analysis focusses on the final stage of the iMPL procedure. Thus, one should be concerned about the possible effects of such framing on eliciting risk attitudes if using the original MPL procedure.

incorporate such variations within a single functional form. We pursue the latter path below.

Calibrating for Task Effects

Using the statistical model shown in Table 3, one can "calibrate" for an elicited distribution of risk attitudes that does not have framing or order effects. This entails generating predictions using the estimated model in which these effects are removed. In effect, the statistical model allows one to predict how each subject would have responded if they had faced the symmetric menu and only made one response, conditional on all of their other individual characteristics (such as age, sex, etc.). Figure 5 shows the results of this exercise. The top panel shows the observed distribution, recognizing that this reflects predictions "observed" from the estimated statistical model in Table 3. The bottom panel shows the adjusted or calibrated distribution, which is generally less risk averse.

The constructive implication is that the existence of framing and task order effects can be taken into account when drawing inferences from elicitation procedures.

Joint Effects

Now consider the effects of key demographic variables considered jointly. To do so we generate predictions of the CRRA from the statistical model underlying Table 3, calibrating for task effects, and simply stratify these predictions according to the demographic variable of interest. For example, the men in our sample have a number of characteristics that differ from the women apart from sex: they tend to be younger, have a higher income, live in Copenhagen, and are more likely to be employed, a student, skilled with some post-secondary education, and have higher education. Several of these characteristics had a significant marginal effect on risk attitudes, hence it is possible that the joint effect of sex along with the characteristics correlated with it could have a significant effect on risk attitudes.

We find that sex still has no effect on risk aversion.[27] Being single is associated with a higher aversion to risk, which suggests numerous jokes about the uncertainties of married life that we will avoid. The effect is small (+0.08 in terms of CRRA), but statistically significant using a *t*-test. Owners of apartments or homes have a lower aversion to risk (-0.10), despite the fact that the marginal effect of ownership was negligible. Retired individuals have much lower aversion to risk (-0.19), in contrast to the absence of a significant marginal effect. The joint effect on risk aversion of being a student is even larger than the marginal effect (+0.45 *versus* +0.33), whereas the joint effect of being skilled with some post-secondary education is smaller (+0.25 *versus* +0.44). The joint effect of having substantial higher education is only half of the marginal effect, but is statistically significant. The joint effect of income levels is roughly the same as the marginal effect, but is statistically significant. Thus lower income individuals have lower aversion to risk (-0.07), whereas higher income individuals have greater aversion to risk (+0.06), each in relation to middle-income individuals. This result is consistent with income being correlated with many of the other characteristics included in our analysis, as might be expected. Living in greater Copenhagen is associated with a significant joint effect, and is associated with higher risk aversion (+0.12). The same effect is associated with living in a larger city, although it is slightly smaller (+0.08).

C. *Constancy of Relative Risk Aversion*

The CRRA characterization of risk attitudes is popular in theoretical and applied work, no doubt due to it's tractability. If the analyst has the flexibility of allowing for other functional forms,[28]

---

[27] To better compare to previous studies, which sometimes only condition on sex, we also estimated the interval regression model controlling only for sex and found no effects. Adding experimenter and task effects did not change this result.

[28] That is, it is entirely appropriate to maintain the CRRA assumption if one requires such structure in order to generate predictions in some context. For example, there is a relatively elaborate theory of bidding behavior in first-price auctions that relies on such representations of risk attitudes in order to solve for closed-form Bayesian Nash equilibria. If the CRRA characterization is rejected for some income domains, then this should be taken into account when evaluating tests of those bidding theories that treat CRRA as a maintained assumption. Ideally one would go back and re-formulate the theory to accommodate such a result, but that may not always be feasible. Moreover, even if CRRA is not globally valid over a given income domain, it may still be locally valid of a subset of that domain.

is it an appropriate functional form for the domain of incomes considered in our experiments? We answer this question by characterizing our data with a more flexible functional form that nests CRRA, the EP function introduced earlier, and directly testing for constancy of RRA.

We assume initially that we can empirically characterize the distribution of risk attitudes for all subjects using one functional form, and assume away any heterogeneity across individuals. This might seem like an odd thing to do, apart from the fact that it has been a popular thing to do.[29] Once we have evaluated the CRRA characterization in this setting, we consider the effects of allowing heterogeneity in the EP characterization. Conditioning on individual heterogeneity makes a significant difference to the characterization of risk attitudes, hence the need to proceed in these steps.

Maximum likelihood estimates of the EP model can be used to generate the RRA for different income levels.[30] Figure 6 displays the results, including a 95% confidence interval. The estimates are very precise, even if they indicate large changes in the value of RRA over this income domain. Using this characterization of utility, subjects behave as if risk-loving for the lower income levels, up to around 1,250 DKK, and then become significantly risk averse. Their RRA is proximately constant for income levels beyond 2,000 DK.

The most remarkable feature of Figure 6 is obviously the change from risk aversion for higher income levels to risk loving for lower income levels. We did not conduct any experiments that confronted subjects with a loss frame, so our analysis is restricted entirely to the gain frame.[31] This pattern is strikingly consistent with the "reflection effect," proposed by Kahneman and Tversky

---

[29] See Camerer and Ho [1994; p.186-7] for a defence of this approach. They clearly propose it as an exploratory approach to complement more elaborate analyses that allow for individual heterogeneity.

[30] The likelihood function employs a probit function linking observed choices to the probability of that choice conditional on parameter values.

[31] The term "loss frame" refers to a task in which the choices are framed as losses from some reference point. All of the non-hypothetical experiments that we know of provide a positive reference point such that the subject cannot suffer a net loss (e.g., Battalio, Kagel and Jiranyakul [1990] and Kagel, MacDonald and Battalio [1990]). An experiment in which the subject faces choices in which some outcomes involved net losses should be referred to as involving choices in the *loss domain*. These are important semantics when evaluating the experimental evidence accurately, since none of the non-hypothetical evidence pertains to the loss domain.

[1979], at least in its most general form. The examples they provide, and their intuition, all refer to risk loving behavior over losses and risk averse behavior over gains, but they are careful to always refer to losses and gains relative to some "reference point" (e.g., p.279). Implicit in their hypothetical experiments and intuition is a reference point of zero,[32] but that is not essential. If one modifies the notion of a reference point to be some subjective income level below which the subject is "willing to gamble," then it need not be zero for all subjects. If non-experimental income is reliable and large enough, a subject may well be willing to "take a flutter" for low income levels. In any event, this pattern is observed clearly in Figure 6. This pattern is, in fact, exactly the one proposed by Markowitz [1952], who Kahneman and Tversky [1979; p.268] cite as a precursor.

On the other hand, Figure 6 is simply an empirical version of a utility function that is well-defined from the perspective of EUT. If one had selected income domains from Figure 6 such that some experiments involved tasks where subjects were risk loving, and other experiments involved tasks in income domains where subjects were risk averse, and then applied a CRRA specification, it might appear that there had been a shift in risk preferences. But this is simply an artefact of applying a CRRA specification over an income domain for which CRRA is not *globally* valid, which is the fundamental lesson from Holt and Laury [2002]. Thus we see a relatively simple resolution of the apparent paradox of the "reflection effect" and EUT.[33]

It is a simple matter to extend the statistical analysis to allow for heterogeneity. Figure 7 reports the results of a maximum-likelihood estimation in which the r parameter of the EP utility function depends on the same individual and task characteristics as the CRRA regression analysis in

---

[32] Or, more accurately, wealth prior to the choice, since they explicitly view the utility/value function as defined over gains and losses in income relative to that reference point. In their examples they always view the subject as facing choices in the loss domain, at least as far as experimental income is concerned. But those examples are clearly specific cases of their general version. Our results suggest that the reflection effect may occur at positive income levels.

[33] Again, we have no data on loss frame choices, but it is reasonable to extrapolate the qualitative pattern of Figure 6 for low income levels. One could in fact undertake such an extrapolation numerically, but that is not a convincing substitute for experiments in the loss frame. Recent experiments by Laury and Holt [2002] and Harbaugh, Krause and Vesterlund [2002] suggest a significant reduction or elimination of the reflection effect around zero when one moves from hypothetical tasks to non-hypothetical tasks.

Table 3. To see the effect of allowing for demographics, we also show the corresponding results when no individual heterogeneity is allowed for. In both cases in Figure 7 we restrict the sample to those subjects facing the symmetric frame, to focus on essentials.

The result of allowing for individual heterogeneity is a significant difference in the overall characterization of risk attitudes. Average RRA is roughly constant over the income ranges considered here, in contrast to the large changes inferred when assuming homogeneity. In particular, there is no evidence of significant risk loving behavior at low stake levels. However, there is a dramatic increase in uncertainty about risk attitudes at low stake levels, such that one cannot rule out risk neutrality or risk loving behavior. To some extent this is just due to sampling uncertainty when one evaluates an estimated statistical relationship away from the sample mean stake level, but it also reflects diverse risk attitudes by subjects differentiated by observable characteristics. Moreover, the plateau of constant RRA is much lower than when heterogeneity is assumed away. With no heterogeneity, the characterization indicated extreme risk aversion, with average RRA values close to 2 for medium to higher stake levels. When heterogeneity is allowed for, on the other hand, this plateau is much more reasonable, and average RRA is well below 1.

Why is there more uncertainty about risk attitudes when one controls for individual heterogeneity? The reason for this uncertainty is that several of the individual characteristics are statistically significant, as we found with the CRRA regression analysis. Thus the EP function for some individuals is significantly different from the EP function for other individuals, so when one evaluates the RRA for the average individual these differences imply that the average RRA is less representative of the sample of EP functions. Actually, this "average individual" is literally a counterfactual, synthetic individual whose characteristics are equal to the average for the sample. For example, this synthetic individual is 51% female (Table 2), whereas the actual individuals were either female or they were not. If sex had been a significant determinant of risk attitudes, this would have been one of the factors causing wider confidence intervals in the right panel of Figure 7.

D. Variants from the Standard Model

We consider the robustness of our main findings to a popular variant of the standard expected utility theory model: probability weighting.[34] Expected utility is defined in standard theory by the probability-weighted utilities of the final prizes of the lotteries the subject makes choices over. In standard theory expected utility is just a weighted average, where the weights are the probabilities of the lotteries. Although these are the subjective probabilities, it is reasonable for tasks as simple as ours to assume that the subjective probabilities are the objective probabilities. An alternative approach, popularized in the earlier variants of Prospect Theory and due to Edwards [1962], is to allow the probability weights to be some non-linear function of the objective probabilities. The original idea, at least as expressed in Kahneman and Tversky [1979; p.274ff.], was that this reflected some psychological "editing phase" in which the subjects generated a mental representation of the task.

For a specific functional form, we employed the power function in the gain domain used by Tversky and Kahneman [1992] and others.[35] Thus the decision weights used instead of probabilities are the probabilities raised to some power, where that power is to be estimated as part of the overall maximum-likelihood exercise that also estimates the parameters of the EP utility function. The results, shown in the right-hand panel of Figure 8, indicate that there is a significant change in the EP representation: uncertainty about risk attitudes is increased, and although the best estimates for RRA indicate approximate CRRA, the standard errors are very large. These results are consistent with the fact that the characterization of risk interacts with the characterization of probabilities, such that it is not possible to be very precise about one without imposing some constraints on the other.

[34] An alternative approach to relaxing the predictions of EUT has been to allow for there to be some stochastic deviation from the prediction of the theory and observed choice. If one assumes homogeneous risk attitudes, for example, some such assumption is necessary to account for the blunt fact that no two subjects made the same choices (e.g., see Camerer and Ho [1994; p.186] and HL [2002; p.1652]). Of course, simply allowing subjects to have different preferences will mitigate the need for this specification. This assumption has nothing to commend it in theory, and while it is parsimonious that is hardly an advantage when one considers the need to augment the statistical analysis with ill-motivated noise parameters (as argued eloquently by Ballinger and Wilcox [1997]). Appendix B discusses this methodological perspective in more detail.

[35] Wu and Gonzalez [1996] and Prelec [1998] discuss more general functional forms.

This finding does not invalidate EUT or the probability-weighting variant, but just serves as a reminder that allowing flexibility in both appears to "over-determine" these data.

4. Conclusions

We demonstrate that it is possible to elicit attitudes to risk from individuals in the field using real economic commitments, and that those attitudes are in an *a priori* plausible range. There are variations in risk attitudes across significant socio-demographic characteristics of the Danish population, implying that welfare evaluations of government policies for those individuals should take these differences into account.[36] We also show that average risk aversion is roughly constant with respect to the income levels considered here, although there is considerable uncertainty about specific risk attitudes for "low" income levels. Since those small amounts are in the range of most laboratory experiments, they imply that experimenters should be concerned about the sensitivity of the values of risk aversion their subjects might exhibit. On the other, if the approximate constancy of relative risk aversion for "larger" amounts of income extends beyond the domain of the prizes used in our experiments, our results imply that CRRA may be a useful characterization for policy purposes.

Our results consistently support the need to recognize the heterogeneity of risk attitudes across individual subjects. If one is trying to characterize risk attitudes for our sample as a scalar value of relative risk aversion, the confidence intervals would be quite wide. Even the choice of appropriate functional form, and the assumption of constancy of relative risk aversion, depends on

---

[36] Most welfare analyses of government policy are associated with the household, or representative household types, due to data limitations on economic activities of individuals within the household (e.g., see Harrison, Jensen, Lau and Rutherford [2001] or Harrison, Rutherford and Tarr [2003]). Our results for individuals can be used to generate weighted averages for households composed of individuals with different characteristics. However, we caution that risk attitudes for households as decision-making units might very well differ from simple weighted averages of the risk attitudes of the members of that household. Such attitudes could be elicited directly using simple variants of the procedures used here, and that is an important area for future research. But it is not obvious from an individualistic welfare perspective whether it is methodologically better to use household estimates of risk aversion or individual estimates of risk aversion.

whether one imposes homogenous risk preferences across individuals. But this sensitivity to specification is just a reflection of the fact that risk attitudes for individuals differ. This result has important implications for the characterization of risk attitudes in policy applications, theoretical modeling, and experimental economics.

Table 1: Payoff Matrix in the Holt and Laury Risk Aversion Experiments

Default payoff matrix for scale 1

Lottery A				Lottery B				EV^A	EV^B	Difference	Open CRRA Interval if Subject Switches to Lottery B
p($2)		p($1.60)		p($3.85)		p($0.10)					
0.1	$2	0.9	$1.60	0.1	$3.85	0.9	$0.10	$1.64	$0.48	$1.17	-∞, -1.71
0.2	$2	0.8	$1.60	0.2	$3.85	0.8	$0.10	$1.68	$0.85	$0.83	-1.71, -0.95
0.3	$2	0.7	$1.60	0.3	$3.85	0.7	$0.10	$1.72	$1.23	$0.49	-0.95, -0.49
0.4	$2	0.6	$1.60	0.4	$3.85	0.6	$0.10	$1.76	$1.60	$0.16	-0.49, -0.15
0.5	$2	0.5	$1.60	0.5	$3.85	0.5	$0.10	$1.80	$1.98	-$0.17	-0.15, 0.14
0.6	$2	0.4	$1.60	0.6	$3.85	0.4	$0.10	$1.84	$2.35	-$0.51	0.14, 0.41
0.7	$2	0.3	$1.60	0.7	$3.85	0.3	$0.10	$1.88	$2.73	-$0.84	0.41, 0.68
0.8	$2	0.2	$1.60	0.8	$3.85	0.2	$0.10	$1.92	$3.10	-$1.18	0.68, 0.97
0.9	$2	0.1	$1.60	0.9	$3.85	0.1	$0.10	$1.96	$3.48	-$1.52	0.97, 1.37
1	$2	0	$1.60	1	$3.85	0	$0.10	$2.00	$3.85	-$1.85	1.37, ∞

Note: The last four columns in this table, showing the expected values of the lotteries and the implied CRRA intervals, were not shown to subjects.

Figure 1: First Level of the iMPL Elicitation Format

Decision	Option A	Option B	Choice
1	2000 kr. if number of ball is 1-10 1600 kr. if number of ball is 11-100	3850 kr. if number of ball is 1-10 100 kr. if number of ball is 11-100	◉ A ○ I ○ B
2	2000 kr. if number of ball is 1-20 1600 kr. if number of ball is 21-100	3850 kr. if number of ball is 1-20 100 kr. if number of ball is 21-100	◉ A ○ I ○ B
3	2000 kr. if number of ball is 1-30 1600 kr. if number of ball is 31-100	3850 kr. if number of ball is 1-30 100 kr. if number of ball is 31-100	◉ A ○ I ○ B
4	2000 kr. if number of ball is 1-40 1600 kr. if number of ball is 41-100	3850 kr. if number of ball is 1-40 100 kr. if number of ball is 41-100	○ A ○ I ◉ B
5	2000 kr. if number of ball is 1-50 1600 kr. if number of ball is 51-100	3850 kr. if number of ball is 1-50 100 kr. if number of ball is 51-100	○ A ○ I ◉ B
6	2000 kr. if number of ball is 1-60 1600 kr. if number of ball is 61-100	3850 kr. if number of ball is 1-60 100 kr. if number of ball is 61-100	○ A ○ I ◉ B
7	2000 kr. if number of ball is 1-70 1600 kr. if number of ball is 71-100	3850 kr. if number of ball is 1-70 100 kr. if number of ball is 71-100	○ A ○ I ◉ B
8	2000 kr. if number of ball is 1-80 1600 kr. if number of ball is 81-100	3850 kr. if number of ball is 1-80 100 kr. if number of ball is 81-100	○ A ○ I ◉ B
9	2000 kr. if number of ball is 1-90 1600 kr. if number of ball is 91-100	3850 kr. if number of ball is 1-90 100 kr. if number of ball is 91-100	○ A ○ I ◉ B
10	2000 kr. if number of ball is 1-100	3850 kr. if number of ball is 1-100	○ A ○ I ◉ B

OK

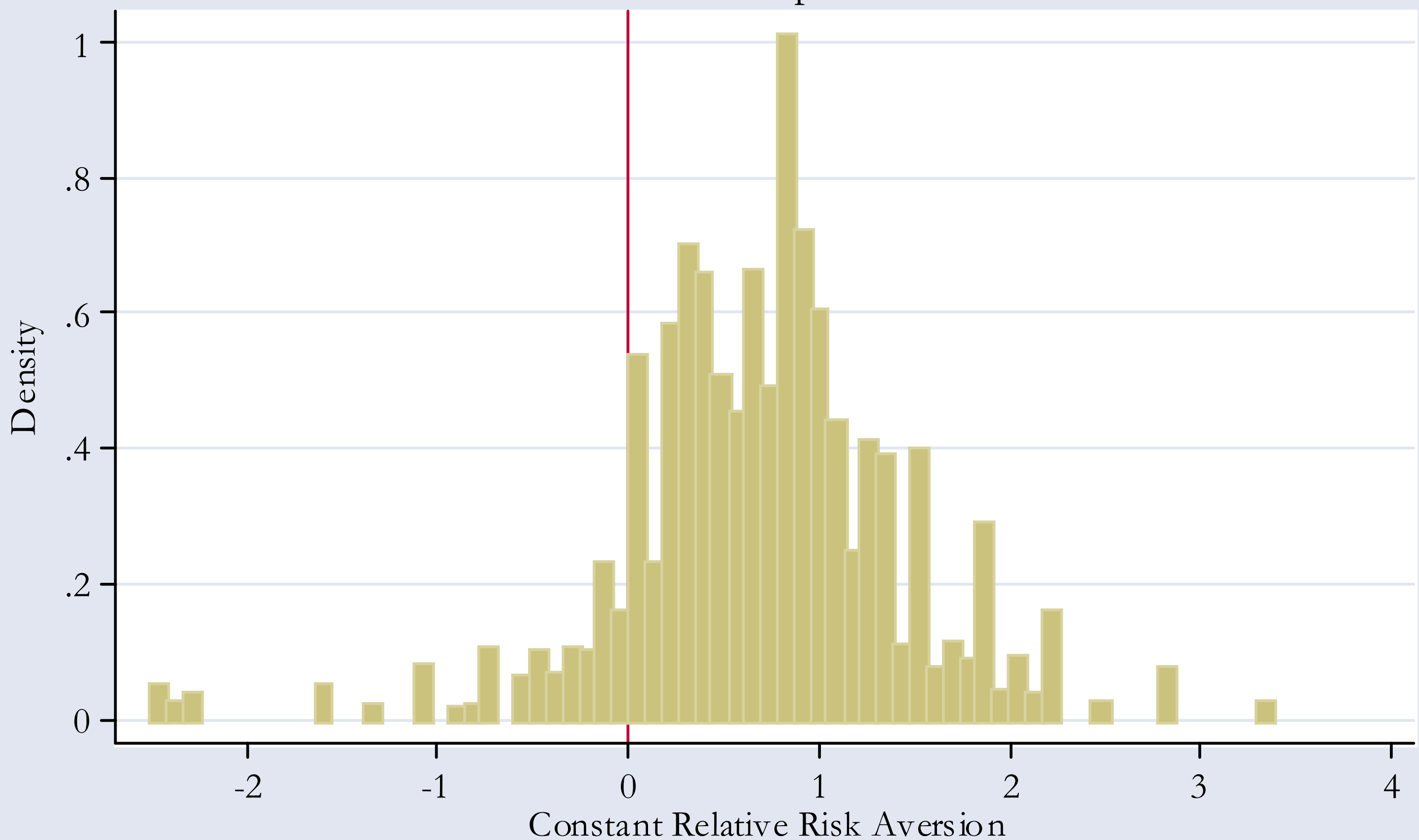
Figure 2: Second Level of the iMPL Elicitation Format

Decision	Option A	Option B	Choice
1	2000 kr. if number of ball is 1-30 1600 kr. if number of ball is 31-100	3850 kr. if number of ball is 1-30 100 kr. if number of ball is 31-100	A I B
2	2000 kr. if number of ball is 1-31 1600 kr. if number of ball is 32-100	3850 kr. if number of ball is 1-31 100 kr. if number of ball is 32-100	A I B
3	2000 kr. if number of ball is 1-32 1600 kr. if number of ball is 33-100	3850 kr. if number of ball is 1-32 100 kr. if number of ball is 33-100	A I B
4	2000 kr. if number of ball is 1-33 1600 kr. if number of ball is 34-100	3850 kr. if number of ball is 1-33 100 kr. if number of ball is 34-100	A I B
5	2000 kr. if number of ball is 1-34 1600 kr. if number of ball is 35-100	3850 kr. if number of ball is 1-34 100 kr. if number of ball is 35-100	A I B
6	2000 kr. if number of ball is 1-35 1600 kr. if number of ball is 36-100	3850 kr. if number of ball is 1-35 100 kr. if number of ball is 36-100	A I B
7	2000 kr. if number of ball is 1-36 1600 kr. if number of ball is 37-100	3850 kr. if number of ball is 1-36 100 kr. if number of ball is 37-100	A I B
8	2000 kr. if number of ball is 1-37 1600 kr. if number of ball is 38-100	3850 kr. if number of ball is 1-37 100 kr. if number of ball is 38-100	A I B
9	2000 kr. if number of ball is 1-38 1600 kr. if number of ball is 39-100	3850 kr. if number of ball is 1-38 100 kr. if number of ball is 39-100	A I B
10	2000 kr. if number of ball is 1-39 1600 kr. if number of ball is 40-100	3850 kr. if number of ball is 1-39 100 kr. if number of ball is 40-100	A I B
11	2000 kr. if number of ball is 1-40 1600 kr. if number of ball is 41-100	3850 kr. if number of ball is 1-40 100 kr. if number of ball is 41-100	A I B

OK

Figure 3: Distribution of CRRA in Denmark
With Symmetric Menu

Mid-Point of Raw Responses from iMPL



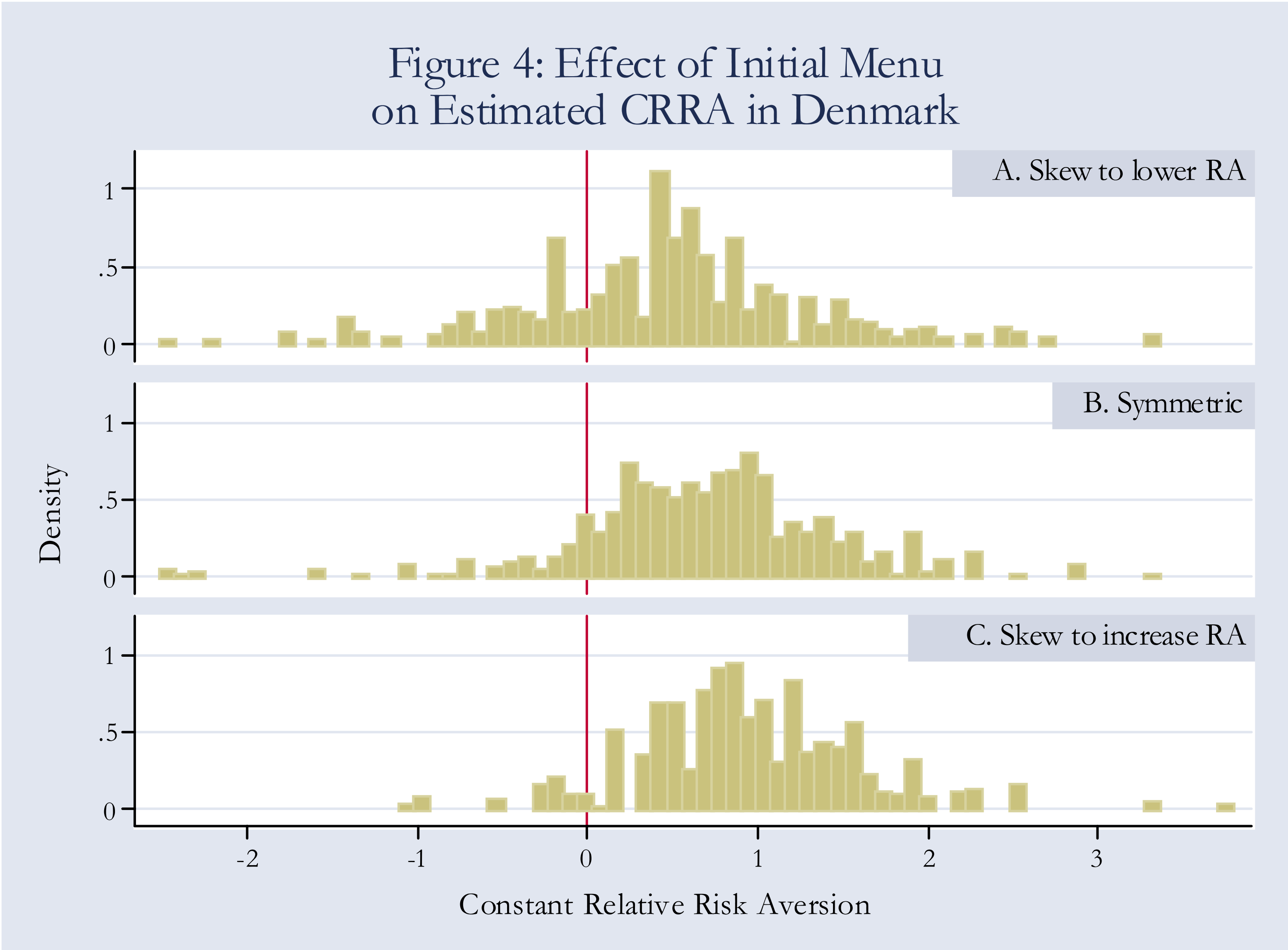


Figure 4: Effect of Initial Menu
on Estimated CRRA in Denmark
A. Skew to lower RA
B. Symmetric
C. Skew to increase RA
Density
1
.5
0
-2
-1
0
1
2
3
Constant Relative Risk Aversion

Table 2: List of Variables and Descriptive Statistics

Variable	Definition	Mean	Standard Deviation
female	Female	0.52	0.50
young	Aged less than 30	0.17	0.38
middle	Aged between 40 and 50	0.28	0.45
old	Aged over 50	0.37	0.48
single	Lives alone	0.20	0.40
kids	Has children	0.29	0.45
nhhd	Number of people in the household	2.5	1.16
owner	Owns own home or apartment	0.69	0.46
retired	Retired	0.16	0.36
student	Student	0.092	0.29
skilled	Some post-secondary education	0.51	0.50
longedu	Substantial higher education	0.37	0.48
IncLow	Lower level income	0.34	0.48
IncHigh	Higher level income	0.34	0.47
copen	Lives in greater Copenhagen area	0.27	0.44
city	Lives in larger city of 20,000 or more	0.39	0.49
experimenter	Experimenter Anderson (default is Lau)	0.55	0.53

Table 3: Statistical Model of Risk Aversion Responses

Random-effects interval regression,
with the CRRA interval chosen by the subject as the dependent variable.

N=934, based on 245 subjects.

Variable	Description	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		0.28	0.28	0.32	-0.28	0.83
skewLO	Skew towards risk loving	-0.004	0.11	0.97	-0.21	0.21
skewHI	Skew towards risk aversion	0.33	0.11	0.00	0.12	0.55
Task2	Second risk task	0.32	0.06	0.00	0.21	0.44
Task3	Third risk task	0.17	0.06	0.00	0.06	0.29
Task4	Fourth risk task	0.13	0.06	0.02	0.02	0.25
experimenter	Experimenter effect	-0.07	0.09	0.45	-0.25	0.11
female	Female	-0.05	0.09	0.57	-0.23	0.13
young	Aged less than 30	0.21	0.18	0.24	-0.14	0.55
middle	Aged between 40 and 50	-0.30	0.14	0.03	-0.57	-0.03
old	Aged over 50	-0.13	0.16	0.41	-0.45	0.18
single	Lives alone	0.08	0.15	0.57	-0.20	0.37
kids	Has children	0.02	0.14	0.87	-0.25	0.30
nhhd	Number in household	0.03	0.06	0.63	-0.09	0.15
owner	Own home or apartment	0.08	0.12	0.54	-0.17	0.32
retired	Retired	0.09	0.15	0.55	-0.20	0.38
student	Student	0.33	0.17	0.05	-0.01	0.67
skilled	Some post-secondary education	0.45	0.14	0.00	0.18	0.72
longedu	Substantial higher education	-0.24	0.14	0.09	-0.52	0.04
IncLow	Lower level income	-0.09	0.12	0.47	-0.33	0.15
IncHigh	Higher level income	0.04	0.12	0.71	-0.19	0.28
copen	Lives in Copenhagen area	0.13	0.12	0.28	-0.11	0.37
city	Lives in larger city of 20,000 or more	0.10	0.11	0.38	-0.12	0.31
sigma_u	Standard deviation of random individual effect	0.59	0.04	0.00	0.52	0.66
sigma_e	Standard deviation of residual	0.62	0.02	0.00	0.59	0.66

Notes: Log-likelihood value is -3299.7; Wald test for null hypothesis that all coefficients are zero has a χ^2 value of 81.45 with 22 degrees of freedom, implying a *p*-value less than 0.001; fraction of the total error variance due to random individual effects is estimated to be 0.475, with a standard error of 0.035.

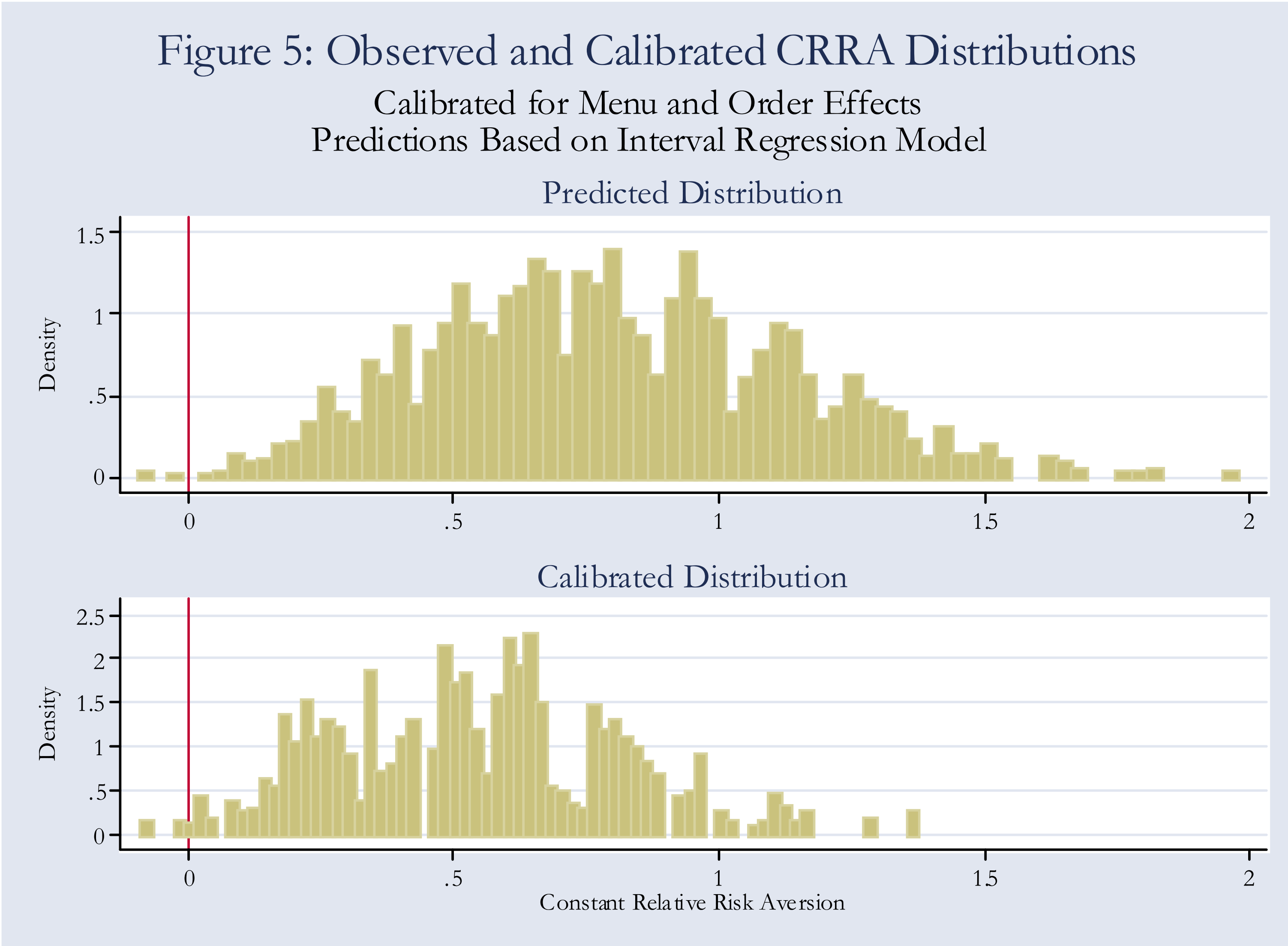


Figure 5: Observed and Calibrated CRRA Distributions
Calibrated for Menu and Order Effects
Predictions Based on Interval Regression Model
Predicted Distribution
Density
1.5
1
.5
0
0
.5
1
1.5
2
Calibrated Distribution
Density
2.5
2
1.5
1
.5
0
0
.5
1
1.5
2
Constant Relative Risk Aversion

Figure 6: Is Relative Risk Aversion Constant?

Maximum likelihood estimates of expo-power utility function
Predicted RRA calibrating for skewness effects
Responses pooled over all four tasks

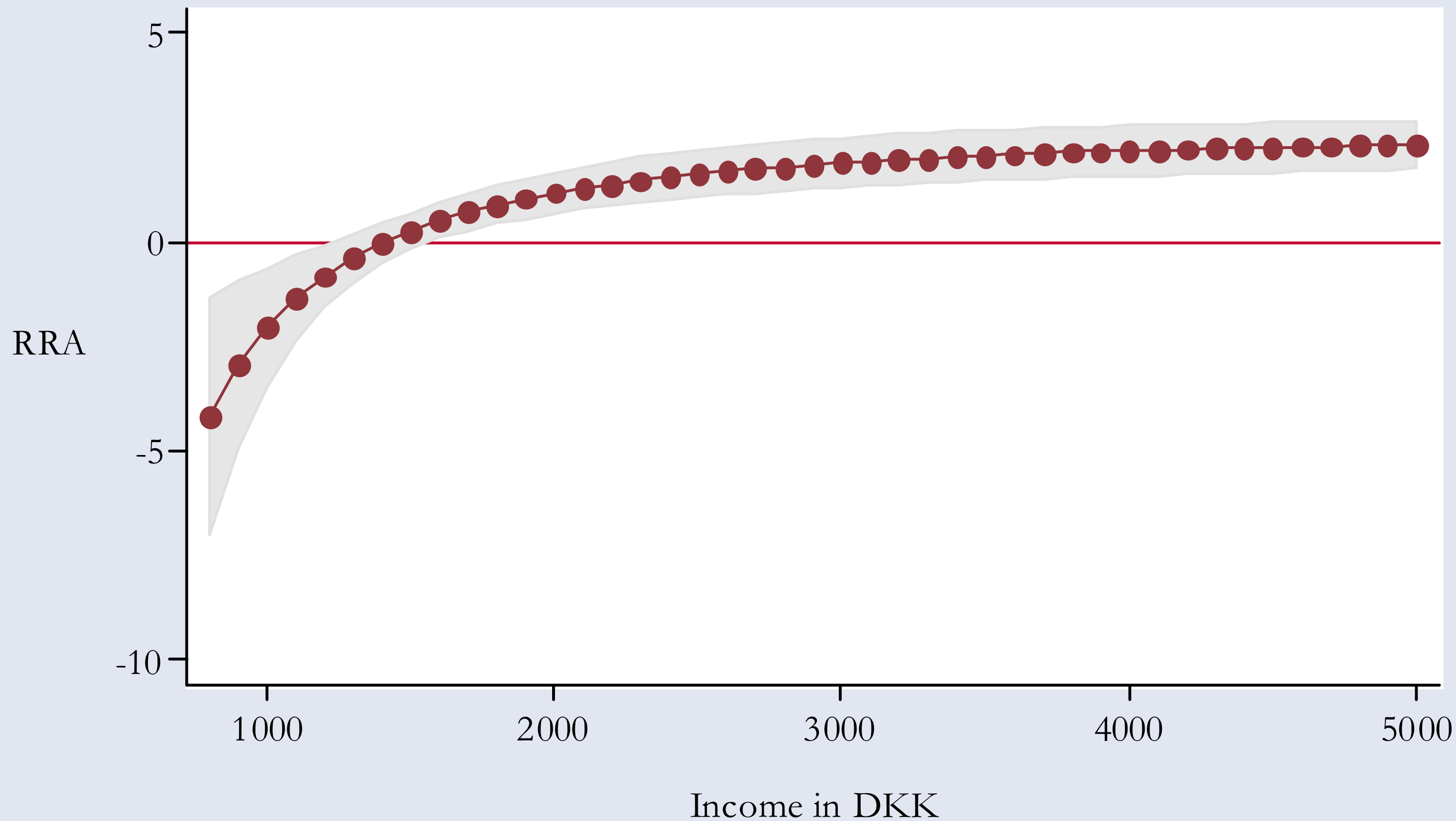


Figure 7: Effect of Demographics on Estimated Risk Attitudes

Maximum likelihood estimates of expo-power utility function

Predicted RRA using sample from symmetric frame

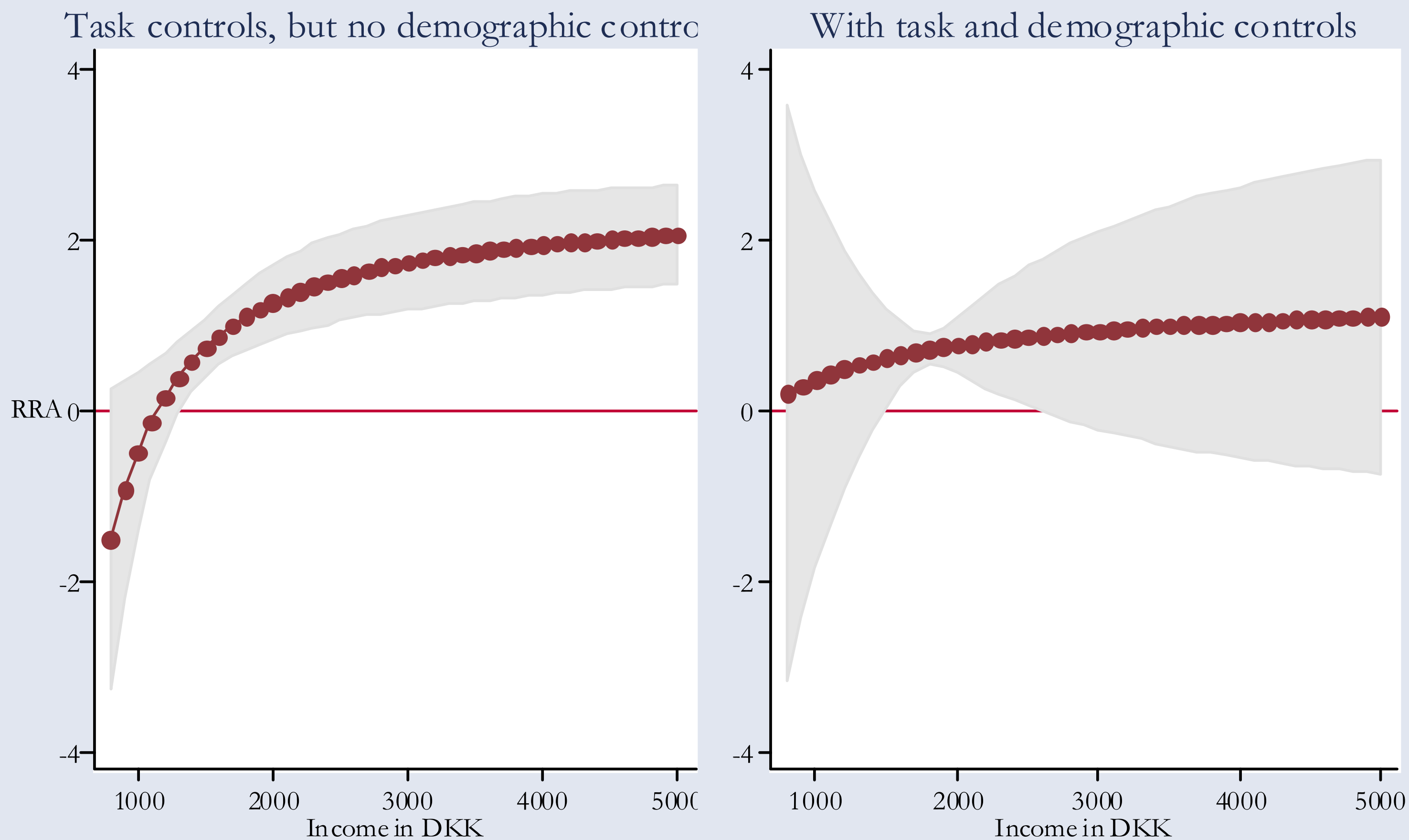
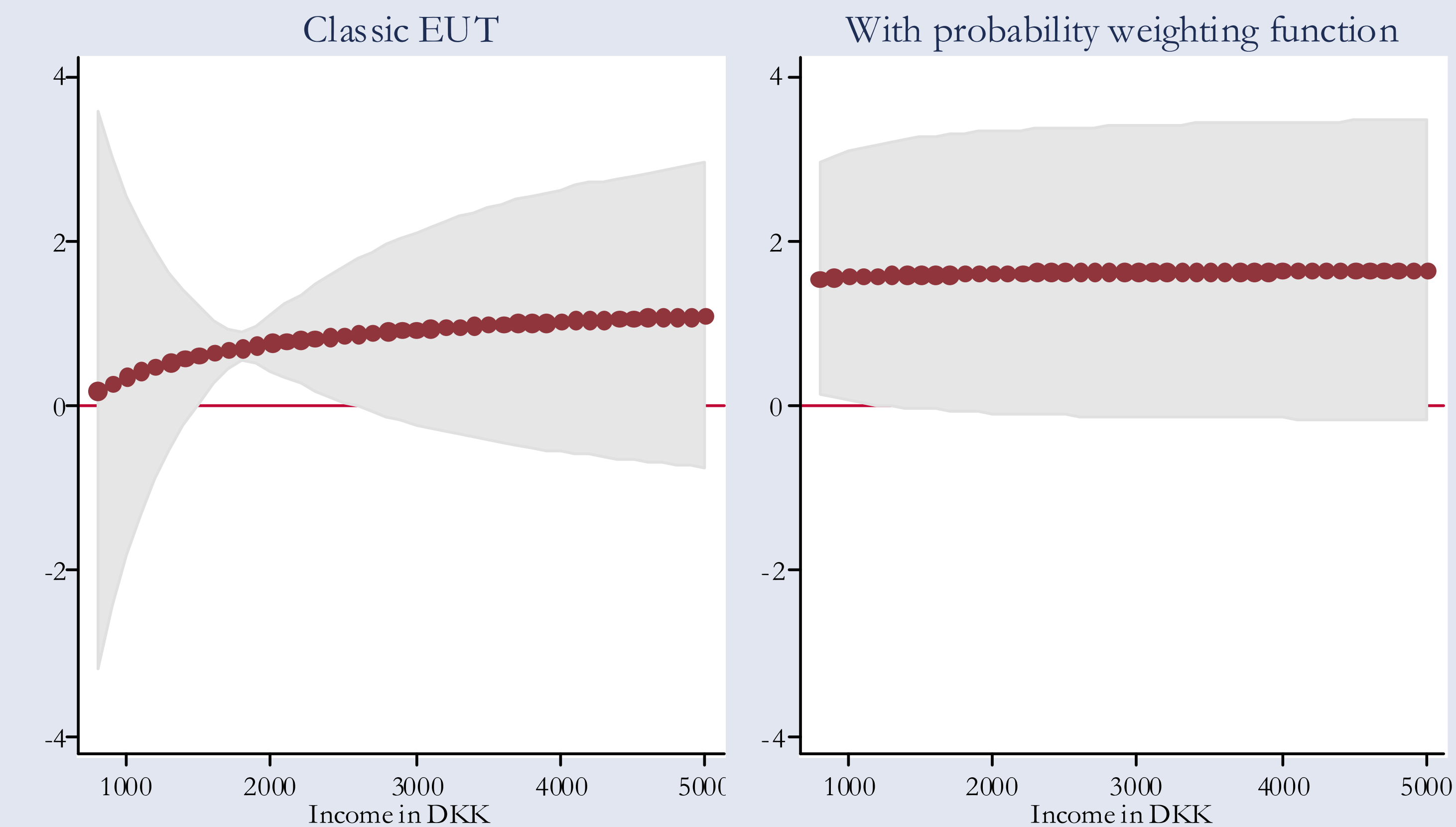


Figure 8: Effect of Probability Weighting on Estimated Risk Attitudes

Maximum likelihood estimates of expo-power utility function
Predicted RRA using sample from symmetric frame
With demographic and task controls



Appendix A: Risk Aversion and Expected Utility Theory

A recent theoretical examination of the role of risk aversion and expected utility theory (EUT) argues that EUT must be rejected for individuals who are risk averse at low monetary stakes. If true, then further tests of EUT are not needed for those individuals who are found to be risk averse in these low stake lottery choices. Rabin [2000] proves a calibration theory showing that if individuals are risk averse over low stakes lotteries then there are absurd implications about the bets those individuals will accept at higher stakes. Rabin [2000] and Rabin and Thaler [2001] allege that this result has general implications for the validity of EUT as a descriptive theory. As explained by Rabin and Thaler [2001; p.222, emphasis added]:

> The logic behind this result is that within the expected utility framework, turning down a moderate stakes gamble means that the marginal utility of money must diminish very quickly. Suppose you have initial wealth of W, and you reject a 50-50 lose \$10/gain \$11 gamble because of diminishing marginal utility of wealth. Then it must be that $U(W + 11) - U(W) \le U(W) - U(W-10)$. Hence, on average you value each of the dollars between W and W + 11 by at most 10/11 as much as you, on average, value each of the dollars between W-10 and W. By concavity, this implies that you value the dollar W + 11 at most 10/11 as much as you value the dollar W-10. Iterating this observation, if you have the same aversion to the lose \$10/gain \$11 bet at wealth level W + 21, *then you value dollar W + 21 + 11 = W + 32 by at most 10/11 as you value dollar W + 21 - 10 = W + 11, which means you value dollar W + 32 by at most 10/11 × 10/11 ≈ 5/6 as much as dollar W-10.* You will value the W + 210th dollar by at most 40 percent as much as dollar W-10, and the W + 900th dollar by at most 2 percent as much as dollar W-10. In words, rejecting the 50-50 lose \$10/gain \$11 gamble implies a 10 percent decline in marginal utility for each \$21 in additional lifetime wealth, meaning that the marginal utility plummets for substantial changes in lifetime wealth. You care less than 2 percent as much about an additional dollar when you are \$900 wealthier than you are now. This rate of deterioration for the value of money is absurdly high, and hence leads to absurd risk aversion.

Thus, a problem for EUT does indeed arise if (a) subjects exhibit risk aversion at low stake levels, *and* (b) one assumes that utility is defined in terms of terminal wealth.[37]

[37] Terminal wealth refers here to the wealth that the subject has prior to coming into the lab plus any income earned in the lab. Watt [2002] and Palacios-Huerta, Serrano and Volij [2002] argue that the degree of

If, on the other hand, one assumes utility is defined over income, this critique will not apply. Consider the step in the argument that is italicized, and which relies critically on the utility function being defined in terms of terminal wealth. If utility were defined in terms of income, then one could not make this step in the argument: all that one could say would be that the person at wealth level W+21 valued dollar W+11 at most 10/11 as much as he valued the dollar W-10. This is the same statement made in the first step of the argument, so there is no basis for making inferences about how the person values much larger stakes.

A careful reading of Rabin [2000] is consistent with this perspective. Consider this passage (p.1288):

> What *does* explain risk aversion over modest stakes? [...] what is empirically the most firmly established feature of risk preferences, *loss aversion*, is a departure from expected-utility theory that provides a direct explanation for modest-scale risk aversion. Loss aversion says that people are significantly more averse to losses relative to the status quo than they are attracted by gains, and more generally that people's utilities are determined by changes in wealth rather than absolute levels.

One can accept the second contention from the above, that subjects use experimental income (i.e., changes in wealth) rather than absolute levels of wealth as the basis for making decisions, independent of the first point about the asymmetry of risk attitudes either side of the status quo.

Whether or not one models utility as a function of terminal wealth (EUTw) or income (EUTi) depends on the setting. Both specifications have been popular. The EUTw specification was widely employed in the seminal papers defining risk aversion and it's application to portfolio choice. The EUTi specification has been widely employed by auction theorists and experimental economists testing EUT, and it is the specification we employ here.

---

relative risk aversion required for Rabin's result are *a priori* implausible. If an individual turned down a small-stakes gamble with a positive expected return for *any* wealth level, including high wealth levels, then that individual must have extremely high relative risk aversion. Hence it could be reasonable for that individual to turns down more generous gambles at higher stakes.

One is tempted to think that this result is well-known since Markowitz [1952] and Samuelson [1952; ¶13, p.676], but that may just be a hindsight bias. Cox and Sadiraj [2003] and Rubinstein [2002] make these points quite clearly. Cox and Sadiraj [2003] go further to propose a generalization of EUTw and EUTi that allows initial wealth to be an argument of the utility function along with income (as long as initial wealth is not simply added to income, which would be EUTw). They also note that "loss aversion," the alternative favored by Rabin [2000] and Rabin and Thaler [2001] as a descriptive model of low-stakes risk aversion, is perfectly consistent with EUTi.

Rubinstein [2002] draws the important connection between adopting an EUTi assumption and the question of temporal consistency of preferences, since the income that one received in today's experiment must be "integrated" in some consistent way with the income received in the past (viz., wealth prior to the experiment). This suggests links back to the older literature on the "asset integration hypothesis," reviewed in this context by Quizon, Binswanger and Machina [1984]. In other words, just because one adopts an EUTi characterization and thereby avoid the problems posed by Rabin [2000], one is not free to make any arbitrary assumptions about behavior over time. The laboratory evidence on this matter has it's own controversies: see Frederick, Loewenstein and O'Donoghue [2002] and Coller, Harrison and Rutström [2003].

Appendix B: Representative Decision Maker Models

It has been common in empirical work in risk attitudes and expected utility theory to assume a representative decision maker.[38] Two significant exceptions to this assumption include Hey and Orme [1994] and Ballinger and Wilcox [1997]. This assumption amounts to starting with the assumption that each subject has the same risk attitudes. Since this assumption does "obvious"

[38] For example, see Camerer and Ho [1994; p.186] and HL [2002; p.1652].

violence to the data when the two are put in the same room, the analyst immediately grabs for one or other assumption about the "error structure" to account for the data. Within these error structures are implicit, reduced form models that allow individuals to have differences in preferences or choices, but these are below the surface.

Since we eschew the representative decision maker assumption from the outset, and it makes a significant difference to the implied characterization of risk attitudes, it is useful to review one recent example of its use by Holt and Laury [2002].

Figure A1 shows the main responses in the HL experiments. Consider the top left panel, which shows the average number of choices of the "safe" option A in each problem. Thus in problem 1, which is row 1 in Table 1, virtually everyone chooses option A (the safe choice). By the time the subjects get to problem 10, which is the last row in Table 1, virtually everyone has switched over to problem B, the "risky" option. The dashed line marked RN shows the prediction if each and every subject were risk-neutral: in this case everyone would choose option A up to problem 4, then everyone would choose option B thereafter. The solid line marked 1X shows the observed behavior in task #1, the low-payoff case in the design of HL. The solid line marked 20X shows the observed behavior in their task #3, the high-payoff case in which the base payoffs were scaled up by 20. The top right panel in Figure A1 shows comparable data for the 50x problems, and the bottom left panel shows comparable data for the 90x problems.[39]

We examine the bottom-right panel later.

HL proceed with their analysis by looking at the first three pictures and drawing two

[39] The control data in these three panels, for the 1x problem, are pooled across all task #1 responses. That is, the task #1 responses in the bottom left panel of Figure A1 are not just the task #1 responses of the individuals facing the 90x problem. Nothing essential hinges on this at this stage of exposition. This makes no difference to the points being made here.

conclusions. First, that one has to introduce some "noise" into any model of the data-generation process, since the observed choices are "smoother" than the risk neutral prediction. A more general way of saying this is to allow subjects to have a specific degree of risk aversion, but to critically assume that they all have exactly the same degree of risk aversion. Thus, if subjects were a little risk averse the line marked RN would shift to the right and drop down a bit to the right, perhaps at problem 6 or 7 instead of problem 5. Figure A2 illustrates one such case with the 1X responses, using the average CRRA of the sample as the basis for the representative decision maker model that assumes some risk aversion. Of course, such an alternative would no longer represent risk-neutral responses, but it would still drop sharply, and that is the point being made by HL when arguing for a noise parameter. Second, and related to the previous explanation, the best-fitting line that maintains the assumption of homogenous risk preferences would have to be a bit to the right of the risk neutral line marked RN. So some degree of risk aversion, they argue, is needed to account for the *location* of the observed averages, quite apart from the need for a noise parameter to account for the *smoothness* of the observed averages.

Both conclusions depend critically on the assumption that every subject in the experiment has the same preferences over risk. The smoothness of the observed averages is easily explained if one allows heterogenous risk attitudes and no noise at all at the individual level: some people drop down at problem 4, some more at problem 5, some more at problem 6, and so on. Indeed, if we allowed each subject to have a CRRA value equal to the mid-point of the interval at which they switched, one could explain the observed averages perfectly – by construction, in fact. The smoothness that the eyeball sees in the observed choices of Figures A1 and A2 is just a counterpart of averaging this heterogeneous process. The fact that *some* degree of risk aversion is needed for *some* subjects is undeniable, from the positive area above the RN line and below the observed choice lines from problems 5 through 10. But it simply does not follow without further statistical analysis that all subjects, or even the typical subject, exhibits significant amounts of risk aversion.

These conclusions follow from inspection of each of the first three panels, and just the RN and observed 1X choice lines in each for that matter. Now turn to the comparison of the low-payoff (1X) and higher-payoff (20X, 50X or 90X) observed choice lines *within* each of the first three panels. The eyeball suggests that the higher-payoff lines are to the right of the 1X lines, which implies that risk aversion increases as the scale of payoffs increases. But this conclusion requires some measures of the uncertainty of these averages. Not surprisingly, the standard deviation in responses is the largest around problems 5 through 7, suggesting that the confidence intervals around these observed choice lines could easily overlap. Again, this is a matter for an appropriate statistical analysis, not eyeball inspection of the averages.

Finally, compare the differences between the 1X and higher-payoff lines as one scans across the first three panels in Figure A1. As the payoff scale gets larger, from 20X to 50X and then to 90X, it appears that the gap widens. That is, if one ignores the issue of standard errors around these averages, it appears that the degree of risk aversion increases. This leads HL to reject CRRA and CARA, and to consider generalized functional forms for utility functions that admit of increasing risk aversion. However, the sample sizes for the 50X and 90X treatments were significantly smaller than those for the 20X treatment: 38 and 36 subjects, respectively, compared to 268 subjects for the 20X treatments. So one would expect that the standard errors around the 50X and 90X high-payoff lines would be much larger than those around the 20X high-payoff lines. This could make it difficult to statistically draw the eyeball conclusion that scale increases risk aversion.

Finally, one needs to account for the fact that all of the high-payoff data in the HL experiments was obtained in a task that followed the low-payoff task. Income effects were controlled for, in an elegant manner (subjects were asked to give up their prior earnings in order to participate in the higher-payoff task). But there could still be *simple order effects* due to experience with the qualitative task. HL recognize the possibility of order effects when discussing why they had the high hypothetical task before the high real task: "Doing the high hypothetical choice task before high real allows us to hold wealth constant and to evaluate the effect of using real incentives. For our purposes, it would not have made sense to do the high real treatment first, since the careful thinking would bias the high hypothetical decisions." The same (correct) logic applies to comparisons of the second real task with the first real task. Indeed, Harrison, Johnson, McInnes and Rutström [2003a] demonstrate that there are order effects in the HL data.

The bottom, right panel examines the data collected by HL in task #1 and task #4, which have the same 1X scale but differ only in terms of the order effect and the accumulated wealth from task #3. These lines appear to be identical, suggesting no order effect, but a closer statistical analysis that conditions on the two differences shows that there is in fact an order effect at work.

The implication is that the only reason that the data have to be smoothed by a "noise parameter" is to correct for the invalidity of the representative decision maker assumption. The fact that there appears to be some need to allow for changes in (relative or absolute) risk aversion as the prize of the lottery changes is an important one, but the evidence here is confounded by order effects. Moreover, the same logic applies in principle to the characterization of choices within a given scale treatment: within the 1X treatment, for example, prizes varied between $0.10 and $3.85. That may not be as quantitatively significant as the 90X scaling, but that is a matter for the data to say. Thus there is no *a priori* or empirical basis for assuming a noise parameter, providing one does not want to insist on the representative decision maker assumption, but some basis for considering non-constant RRA or ARA.

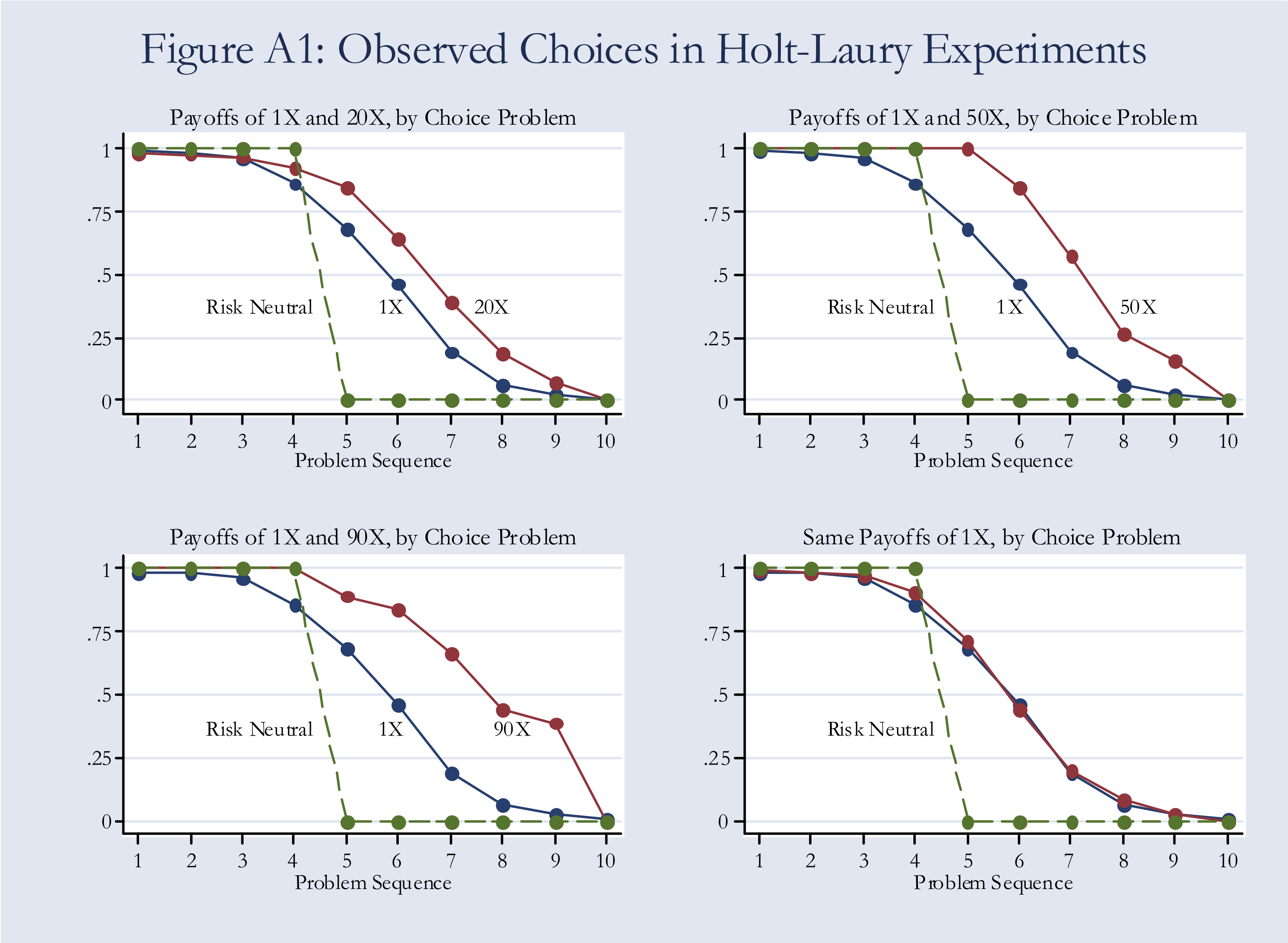
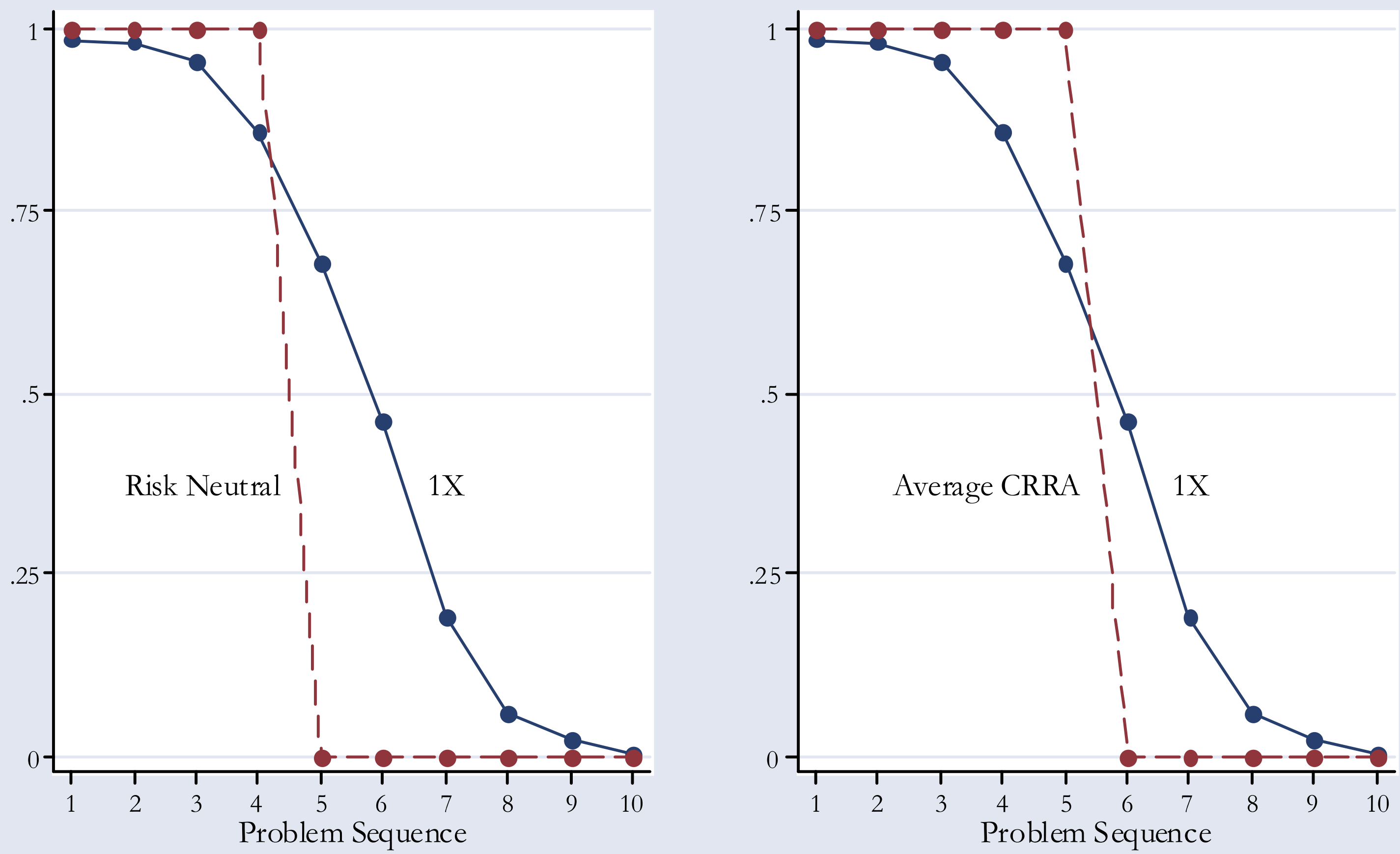


Figure A1: Observed Choices in Holt-Laury Experiments
Payoffs of 1X and 20X, by Choice Problem
Risk Neutral
1X
20X
Problem Sequence
Payoffs of 1X and 50X, by Choice Problem
Risk Neutral
1X
50X
Problem Sequence
Payoffs of 1X and 90X, by Choice Problem
Risk Neutral
1X
90X
Problem Sequence
Same Payoffs of 1X, by Choice Problem
Risk Neutral
Problem Sequence

Figure A2: Observed Choices and Representative Decision Maker Models in the Holt-Laury Experiments

Fraction of Sample Choosing the Safe Option



References

Ballinger, T. Parker, and Wilcox, Nathaniel T., "Decisions, Error and Heterogeneity," *Economic Journal*, 107, July 1997, 1090-1105.

Battalio, Raymond C.; Kagel, John C., and Jiranyakul, K., "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty*, 3, 1990, 25-50.

Camerer, Colin F., and Ho, Teck-Hua, "Violations of the Betweeness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty*, 8, 1994, 167-196.

Coller, Maribeth; Harrison, Glenn W., and Rutström, E. Elisabet, "Are Discount Rates Constant? Reconciling Theory and Observation," *Working Paper 3-31*, Department of Economics, College of Business Administration, University of Central Florida, 2003; available at http://www.bus.ucf.edu/wp/.

Coller, Maribeth, and Williams, Melonie B., "Eliciting Individual Discount Rates," *Experimental Economics*, 2, 1999, 107-127.

Cox, James C., and Sadiraj, Vjollca, "Implications of Small- and Large-Stakes Risk Aversion for Decision Theory," *Unpublished Manuscript*, Department of Economics, University of Arizona, October 2003; available from http://uaeller.eller.arizona.edu/~jcox/papers.html.

Cubitt, Robin P.; Starmer, Chris, and Sugden, Robert, "Dynamic Choice and the Common Ratio Effect: An Experimental Investigation," *Economic Journal*, 108, September 1998, 1362-1380.

Donkers, Bas, and van Soest, Arthur, "Subjective Measures of Household Preferences and Financial Decisions," *Journal of Economic Psychology*, 20(6), 1999, 613-642.

Eckel, Catherine C., and Grossman, Philip J. "Sex and Risk: Experimental Evidence," in C.R. Plott and V.L. Smith (eds.), *Handbook of Results in Experimental Economics* (Amsterdam: North Holland/Elsevier Press, 2003).

Eckel, Catherine; Johnson, Cathleen, and Montmarquette, Claude, "Will the Working Poor Invest in Human Capital? A Laboratory Experiment," *Working Paper 02-01*, Social Research and Demonstration Corporation, February 2002; available from http://www.srdc.org/english/publications/workingpoor.htm.

Edwards, Ward, "Subjective Probabilities Inferred from Decisions," *Psychological Review*, 69, 1962, 109-135.

Frederick, Shane; Loewenstein, George; and O'Donoghue, Ted, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, XL, June 2002, 351-401.

Friedman, David, "Why There Are No Risk Preferrers," *Journal of Political Economy*, 89(3), 1981, 600.

Goeree, Jacob K.; Holt, Charles R., and Palfrey, Thomas R., "Quantal Response Equilibria and Overbidding in Private-Value Auctions," *Journal of Economic Theory*, 104, 2002, 247-272.

Goeree, Jacob K.; Holt, Charles R., and Palfrey, Thomas R., "Risk Averse Behavior in Generalized Matching Pennies Games," *Games and Economic Behavior*, 45(1), October 2003, 97-113.

Grether, David M., and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69(4), September 1979, 623-648.

Harbaugh, William T.; Krause, Kate, and Vesterlund, Lise, "Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses," *Experimental Economics*, 5, 2002, 53-84.

Harrison, Glenn W., Harstad, Ronald M., and Rutström, E. Elisabet, "Experimental Methods and Elicitation of Values," *Experimental Economics*, 2004 forthcoming.

Harrison, Glenn W; Jensen, Jesper; Lau, Morten Igel, and Rutherford, Thomas F., "Policy Reform Without Tears," in A. Fossati and W. Weigard (eds.), *Policy Evaluation With Computable General Equilibrium Models* (New York: Routledge, 2001).

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *Working Paper 3-19*, Department of Economics, College of Business Administration, University of Central Florida, 2003a; available at http://www.bus.ucf.edu/wp/.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Individual Choice and Risk Aversion in the Laboratory: A Reconsideration," *Working Paper 3-18*, Department of Economics, College of Business Administration, University of Central Florida, 2003b; available at http://www.bus.ucf.edu/wp/.

Harrison, Glenn W.; Lau, Morten Igel; 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, 2004 forthcoming); available at http://www.bus.ucf.edu/wp/.

Harrison, Glenn W.; Lau, Morten Igel, and Williams, Melonie B., "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review*, 92(5), December 2002, 1606-1617.

Harrison, Glenn W., and List, John A., "Field Experiments" *Working Paper 3-12*, Department of Economics, College of Business Administration, University of Central Florida, 2003; http://www.bus.ucf.edu/wp/.

Harrison, Glenn W.; Rutherford, Thomas F., and Tarr, David G., "Trade Liberalization, Poverty and Efficient Equity," *Journal of Development Economics*, 71, June 2003, 97-128.

Hey, John D., "Experimental Economics and the Theory of Decision Making Under Uncertainty," *Geneva Papers on Risk and Insurance Theory*, 27(1), June 2002, 5-21.

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), December 2002, 1644-1655.

Jianakoplos, Nancy Ammon, and Bernasek, Alexandra, "Are Women More Risk Averse?", *Economic Inquiry*, 36, October 1998, 620-630.

Kagel, John H.; MacDonald, Don N., and Battalio, Raymond C., "Tests of 'Fanning Out' of Indifference Curves: Results from Animal and Human Experiments," *American Economic Review*, 80(4), September 1990, 912-921.

Kahneman, Daniel; Knetsch, Jack L., and Thaler, Richard H., "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, December 1990, 1325-1348.

Kahneman, Daniel, and Tversky, Amos, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, 47(2), 263-291.

Kapteyn, Arie, and Teppa, Federica, "Hypothetical Intertemporal Consumption Choices," *Economic Journal*, 113, March 2003, C140-C151.

Kirby, Kris N., and Maraković, Nino N., "Delay-discounting probabilistic rewards: Rates decrease as amounts increase," *Psychonomic Bulletin & Review*, 1996, 3(1), 100-104.

Kirby, Kris N.; Petry, Nancy M., and Bickel, Warren K., "Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls," *Journal of Experimental Psychology: General*, 1999, 128(1), 78-87.

Levin, Irwin P.; Snyder, Mary A., and Chapman, Daniel P., "The Interaction of Experiential and Situational Factors and Gender in a Simulated Risky Decision-Making Task," *Journal of Psychology*, 122(2), March 1988,173–181.

Laury, Susan K., and Holt, Charles A., "Further Reflections on Prospect Theory," *Working Paper*, Department of Economics, Georgia State University, 2002; available at http://www.people.virginia.edu/~cah2k/reflect.pdf

Markowitz, Harry, "The Utility of Wealth," *Journal of Political Economy*, 60, April 1952, 151-158.

Mitchell, Robert C., and Carson, Richard T., *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Baltimore: Johns Hopkins Press, 1989).

Murnighan, J. Keith; Roth, Alvin E., and Shoumaker, Francoise, "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty*, 1(1), March 1988, 101-124.

Palacios-Huerta, Ignacio; Serrano, Roberto, and Volij, Oscar, "Rejecting Small Gambles Under Expected Utility," *Working Paper*, Brown University, August 2002; available at http://www.econ.brown.edu/~iph/research.html.

Powell, Melanie, and Ansic, David, "Gender Differences in Risk Behaviour in Financial Decision-Making: An Experimental Analysis," *Journal of Economic Psychology*, 18, 1997, 605-628.

Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66(3), May 1998, 497-527.

Quizon, Jaime B.; Binswanger, Hans P., and Machina, Mark J., "Attitudes Toward Risk: Further Remarks," *Economic Journal*, 94, March 1984, 144-148.

Rabin, Matthew, "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68, 2000, 1281-1292.

Rabin, Matthew and Thaler, Richard, "Anomalies: Risk Aversion," *Journal of Economic Perspectives*, 15, Winter 2001, 219-232.

Robson, Arthur J., "The Evolution of Attitudes to Risk: Lottery Tickets and Relative Wealth," *Games and Economic Behavior*, 14, 1996, 190-207.

Rubin, P. H., and Paul, C. W., "An Evolutionary Model of Taste for Risk," *Economic Inquiry*, 17, 1979, 585–595.

Rubinstein, Ariel, "Comments on the Risk and Time Preferences in Economics," *Unpublished Manuscript*, Department of Economics, Princeton University, 2002; available as http://www.princeton.edu/~ariel/papers/rabin3.pdf

Samuelson, Paul A., "Probability, Utility, and the Independence Axiom," *Econometrica*, 20, 1952, 670-678.

Schubert, Renate; Brown, Martin; Gysler, Matthias, and Brachinger, Hans Wolfgang, "Financial Decision-Making: Are Women Really More Risk Averse?" *American Economic Review Papers & Proceedings*, 89, May 1999, 381-385.

Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 1992, 297-323.

Watt, Richard, "Defending Expected Utility Theory," *Journal of Economic Perspectives*, 16(2), Spring 2002, 227-228.

Wu, George, and Gonzalez, R., "Curvature of the Probability Weighting Function," *Management Science*, 42, 1996, 1676-1690.