Preference Heterogeneity in Experiments: Comparing the Field and Lab

by

Steffen Andersen, Glenn W. Harrison, Morten Igel Lau and E. Elisabet Rutström [†]

December 2004

Abstract. Economists recognize that preferences can differ across individuals. We examine the strengths and weaknesses of lab and field experiments to detect differences in preferences that are associated with standard, observable characteristics of the individual. We consider preferences over risk and time, two fundamental concepts of economics. Our results provide striking evidence that there are good reasons to conduct field experiments. The lab fails to detect preference heterogeneity that is present in the field, obviously due to the demographic homogeneity of the lab. There are also differences in treatment effects measured in the lab and the field that can be traced to interactions between treatment and demographic effects. These can only be detected and controlled for properly in the field data. Thus one cannot simply claim, without additional empirical argument or assumption, that treatment effects estimated in the lab are reliable.

Economists recognize that preferences can differ across individuals. We examine the characterization of preference heterogeneity through controlled experiments. Laboratory experiments have become a generally accepted methodology in mainstream economics, but we are concerned that for some tasks the convenience samples drawn from the relatively homogeneous population on university campuses may lead to misleading inferences. This is particularly likely to be a problem for exercises that entail the explicit elicitation of subjective preferences. If the estimated preference parameters are to be used in policy analysis, the lab findings may be irrelevant since much of the heterogeneity of the broader population is not being represented. We focus on the strengths and weaknesses of lab and field experiments to detect differences in preferences, particularly those differences that are associated with standard, observable characteristics of the individual.

The importance for policy analysis of representing risk and time preferences appropriately, rather than relying on convenience assumptions such as risk neutrality and market interest rates, is demonstrated in Harrison, Lau and Rutström [2004]. They show that a policy-maker could overstate the welfare gains to households by as much as 42% in realistic tax policy simulations by ignoring the risk aversion of the average Dane.[1]

One important finding is that we are able to detect much richer preference heterogeneity when we sample from the field, as compared to convenience samples drawn from college students in our lab experiments. We say "detect" because we do not claim that the heterogeneity is missing in the sample used in the lab, just that the range of variation in characteristics is understandably smaller. The convenience samples are more homogeneous with respect to many observable characteristics, such as age and education level. Thus one would expect that it would be harder to detect differences statistically, even if they were present. Since we have the benefit of comparable lab and field experiments, we can claim that there is preference heterogeneity in the field, and so we know that we are missing something in the lab. We therefore conclude that the lab might not be the

---

[1] For example, assume a policy that is predicted to result in either a zero or a positive (2,000 DKK, for example) effect on the income of the average Danish household, with probabilities ¼ and ¾, respectively. The certainty equivalent of this policy, assuming a risk coefficient in the neighborhood of those we estimate for the field, is about 40% lower than the expected value of 1,500 DKK. Even if we assumed a 90% confidence in the policy prediction instead of 75% confidence, we would be off by almost 20% if we maintained risk neutrality.

best place to search for demographic effects.

Lab findings may even be misleading in cases where individual characteristics do vary in the lab population, such as in the case of the sex of the subject. One might expect that any effect due to sex would be more easily detected in the lab than in the field, since the homogeneity of the lab implicitly controls for any influences on behavior and preferences from other characteristics. Just as lab rats are often bred to be genetically similar so that one can better detect the effects of drug treatments, convenience samples drawn from a relatively homogeneous population presumably allow crisp identification of exogenous effects. Nevertheless, if there are interaction effects between some of the individual characteristics, samples with limited heterogeneity in these characteristics may lead to incorrect inferences. This general point has been made by Botelho, Harrison, Hirsch and Rutström [2005], in the context of inferences about sex and national effects on bargaining behavior.

Similarly, lab findings may be misleading regarding treatment effects, if they interact with demographics that do not vary in the lab. These can only be detected and controlled for properly in the field data. We find, for example, large and significant interaction effects between age and task framing for our risk attitude tests, and also weakly significant interaction effects between sex and framing. Since age does not vary much in the lab, one cannot claim, without additional empirical argument or assumption, that treatment effects estimated in the lab are reliable.

We focus on preference differences in two separate dimensions: aversion to risk and time delay. We use experimental tasks with real monetary rewards to elicit individual risk attitudes and discount rates. These valuation tasks are described in section 1, and build on the risk aversion experiments of Holt and Laury [2002] and the discount rate experiments of Coller and Williams [1999] and Harrison, Lau and Williams [2002].

Our design is implemented in the field in Denmark, to obtain a sample that offers a wider range of individual characteristics than usually found in subject pools recruited at universities.[2] We

---

[2] Our field experiments are "artefactual field experiments" in the terminology of Harrison and List [2004]. One important extension would be to examine if field *surveys* generated reliable estimates of risk aversion and discount rates. Those surveys have the advantage that they can be sent out to much larger samples than we could afford to visit using field experiments. They have the obvious disadvantage that they involve hypothetical tasks, but there is now a substantial literature on the complementary use of (lab) experiments and surveys in which the lab results are used to "statistically

review our field and lab experiments in section 2 and examine the results in section 3.

Our results apply to experiments in which we elicit preferences directly. One might expect such experiments to exhibit more demographic effects than many lab experiments in which preferences over risk and time play a less important role. For example, it is not obvious that our results would cause one to doubt the vast set of experimental results on the relative efficiency of double auction markets relative to posted offer markets. But whenever some treatment effect might depend on risk or time preferences in the lab, then there should be concerns about interaction effects. Risk, in particular, plays a more important role in many experiments than has been widely acknowledged.[3] Furthermore, there are other characteristics besides standard demographics that can vary in the field and dramatically affect behavior, as explained by Harrison and List [2004]: familiarity with the good and familiarity with the task are two of the more obvious factors that are known to be important in the field.

Conversely, it does not follow that "any field experiment" will provide more reliable results that "any lab experiment." It is possible to conduct field experiments with very narrow segments of the population, and indeed desirable to do so for some purposes. It is also possible, although extremely rare, that one brings non-students into the lab to diversify the demographic mix. Hence, we view our findings as shifting the burden of proof onto those who would claim that treatment effects can be reliably estimated in the lab.

1. Valuation Tasks

A. Risk Aversion

We employ a simple experimental measure for risk aversion introduced by Holt and Laury [2002] and extended by Harrison, Lau, Rutström and Sullivan [2005]. Each subject is presented with a choice between two lotteries, which we can call A or B. Table 1 illustrates the basic payoff matrix presented to subjects using lottery prizes from Holt and Laury [2002]. The first row shows that

---

calibrate" the surveys for the effects of hypothetical bias (e.g., Blackburn, Harrison, and Rutström [1994]).
    [3] For example, ultimatum bargaining games, trust or investment games, public goods contribution games, and common pool extraction games, to name an important few.

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 $1.16. As one proceeds down the matrix, the expected value of both

lotteries increases and the expected value of lottery B eventually exceeds 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. 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. In addition

to the A/B choice on each row there is also an option to express indifference, not shown in Table 1.

These data may be analyzed using a constant relative risk aversion (CRRA) characterization

of utility, employing an interval regression model.[4] The CRRA utility of each lottery prize y is

defined as $U(y) = (y^{1-r})/(1-r)$, where r is the CRRA coefficient.[5] The dependent 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 Table 1 one can calculate the implied bounds on the CRRA

coefficient, and these intervals are shown in the final column of Table 1. Thus, for example, a

subject who made 5 safe choices and then switched to the risky alternatives would have revealed a

CRRA interval between 0.14 and 0.41, a subject who made 7 safe choices would have revealed a

---

[4] Holt and Laury [2002] introduce the use of the flexible Expo-Power (EP) utility function, originally developed by Saha [1993], for the analysis of risk aversion in experiments such as these. EP allows varying RRA over the prize income subjects faced in an experiment, thereby generalizing CRRA. Harrison, Lau and Rutström [2004] show that CRRA cannot be rejected as an overall characterization of behavior in the field experiments, but that some demographic segments of the population exhibit behavior consistent with non-CRRA utility functions. Since our concern is with differences in preference heterogeneity across field and lab, we plan to examine our conclusions for robustness to the use of EP in a later draft.
[5] With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(y) = \ln(y)$.

CRRA interval between 0.68 and 0.97, and so on.[6]

The payoffs were in Danish Kroner (DKK) and each task had 4 different prizes, selected so that all 16 prizes span the range of income over which we seek to estimate risk aversion. The four sets of prizes are as follows, with the two prizes for lottery A listed first and the two prizes for lottery B listed next: (A1: 2000 DKK, 1600 DKK; B1: 3850 DKK, 100 DKK), (A2: 2250 DKK, 1500 DKK; B2: 4000 DKK, 500 DKK), (A3: 2000 DKK, 1750 DKK; B3: 4000 DKK, 150 DKK), and (A4: 2500 DKK, 1000 DKK; B4: 4500 DKK, 50 DKK). At the time of the field experiments, the exchange rate was approximately 6.55 DKK per U.S. dollar, so the prizes range from approximately \$7.65 to \$687.

B. Individual Discount Rates

The experimental design for eliciting individual discount rates was introduced in Coller and Williams [1999] and extended by Harrison, Lau and Williams [2002]. The basic question used to elicit individual discount rates is simple: do you prefer \$100 today or \$100+x tomorrow, where x is some positive amount? If the subject prefers the \$100 today then we can infer that the discount rate is higher than x% per day; otherwise, we can infer that it is x% per day or less. The format of our experiments extended this basic question in several ways.

First, a number of such questions were posed to each individual, each question varying x by some amount. When x is zero we would obviously expect the individual to reject the option of waiting for no rate of return. As we increase x we would expect more individuals to take the future income option. For any given individual, the point at which they switch from choosing the current income option to taking the future income option provides a bound on their discount rate. That is, if an individual takes the current income option for all x from 0 to 10, then takes the future income option for all x from 11 up to 100, we can infer that their discount rate lies between 10% and 11%

for this time interval.

Second, we simultaneously pose several questions with varying values of x, and select one question at random for actual payment after all responses have been completed by the individual. In this way the results from one question do not generate income effects which might influence the answers to other questions.

Third, the experiments provided choices between two *future* income options rather than one "instant income" option and one future income option. For example, we offer $100 in one month and $100+$x$ in 7 months, interpreting the revealed discount rate as applying to a time horizon of 6 months. This avoids the potential problem of the subject facing extra transactions costs, including the possibility of default by the experimenter, with the future income option. If the delayed option were to involve greater transactions costs, then the revealed discount rate would include these subjective transactions costs. By having both options entail future income we hold these transactions costs constant.

Subjects were presented with six individual discount rate tasks in the same order, corresponding to six different time horizons: 1 month, 4 months, 6 months, 12 months, 18 months and 24 months. Payoffs to any one subject could range from 3,000 DKK up to 7,697 DKK, and this range converts to $458 and $1.175.

Subjects in the 6-month horizon treatment were given payoff tables illustrated in Table 2. They were told that they must choose between payment Options A and B for each of the 10 payoff alternatives. An option to express indifference was also included, although not illustrated in Table 2. Option A was 3000 Danish kroner (DKK) in all sessions, payable in 1 month. Option B paid 3000 DKK + x DKK in 7 months, where x ranged from annual rates of return of 5% to 50% on the principal of 3000 DKK, compounded quarterly to be consistent with general Danish banking practices on overdraft accounts. The payoff tables provided the annual and annual effective interest rates for each payment option and the experimental instructions defined these terms by way of example.

2. Design of the Field and Lab Experiments

A. Field Experiments

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, essentially following those employed in Harrison, Lau and Williams [2002]:

- 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 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 of 25,000 from the Civil Registry.

- Third, we assigned each county either 1 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 weighted 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. People were told that they would be paid 500 DKK to cover travel costs and that they would have a 10% chance of winning a significant

amount. Response rates were low in some counties, so another 64 invitations were mailed out in these counties to newly drawn subjects. Everyone that gave a positive response was assigned to a session, and our recruited sample was 268.

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.[7] These experiments were conducted in June 2003, and a total of 253 subjects participated in the experiments.[8]

Subjects were first presented with the four separate risk aversion tasks, following a series of training tasks. The tasks were presented in the same order, and differed in the prizes offered. After subjects had completed the four tasks, several random outcomes were generated to determine payments to subjects. One of the four tasks was chosen for each subject, and one of the decision rows in that task was then chosen. When a subject indicated indifference, a random draw determined if the subject received the payments from Lottery A or Lottery B, and another random draw determined if the 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 the final outcome. All payments were made at the end of the experiment. A significant amount of time was spent training subjects on the choice tasks and the randomization procedures.

In the second part of the experiments, subjects were presented with six individual discount rate tasks in the same order, corresponding to the six different time horizons. Future payments were guaranteed by the Danish Ministry of Economic and Business Affairs, and made by automatic transfer from the Ministry's bank account to the subject's bank account. Each subject had a 10%

---

[7] 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 media interviews to stay for two hours; and the fourth person was sent home because they arrived after the instructions had begun and we had already included one unexpected "walk-in" to fill their position.

[8] 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.

chance of receiving the payment associated with his decision in the given choice task and decision row, and we applied the same randomization procedures as those described above.

To allow more refined elicitation of the true preferences, and yet retain the transparency of the incentives of the basic multiple price list, we used a computerized variant on the 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.[9]

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.[10] 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 the first table. If it *is* the row that the subject switched at, another random draw is made to pick a row in the second table that the subject was presented with, and so on.

A natural concern with the MPL is that it might encourage subjects to pick a response in the

_____

[9] 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. Furthermore, the iterative format has some "smarts" built into it: when the values being elicited drop to some specified perceptive threshold (e.g., 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] In fact, one possible explanation of the observation that some subjects switch back and forth between choices in MPL is that they are indifferent. If so, explicitly including an indifference option, as we do here, may be a cleaner way to capture this behavior.

middle of the table, independent of true valuations. There could be a psychological bias towards the middle, although that is not obvious *a priori*. The use of specific values at either end of the table could signal to the subject that the experimenter believes that these are reasonable upper and lower bounds. In some tasks, such as the risk elicitation task, the values are bounded by the laws of probability between 0 and 1, so this is less likely to be a factor compared to the pure psychological anchor of the middle row.

We test for framing effects by varying the cardinal scale of the MPL used, but only in the risk aversion tasks. In the lab experiments we vary the frame in both the risk aversion task and the discount rate 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.[11] As suggested by the treatment names, *skewLO* (*skewHI*) is intended to skew responses to be lower (higher) probabilities if subjects pick in the middle.

B. Lab Experiments

The lab experiments were conducted in October 2003. We recruited 100 subjects from the University of Copenhagen and the Copenhagen Business School. All subjects were recruited using the *ExLab* software.[12] The sessions were announced in 7 different lectures. At each lecture an announcement of the experiment was read aloud, and subjects were asked to enrol for the experiment by accessing *ExLab* through the Danish web page for this project. Of the 100 subjects recruited, 90 showed up for the experiment evenly spread across the 9 sessions. Although several non-students participated, 74 out of the 90 subjects were students. Ages varied from 18 to 32 years, averaging 22.7 years, and only 27% were female.

---

[11] The skewed frames 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.

[12] This recruitment software is available for academic use at http://exlab.bus.ucf.edu. In addition, all instructions for our experiments are provided for public review at the *ExLab* Digital Library at the same location.

We examined the performance of three MPL institutions (MPL, sMPL and iMPL) for each of the three framing conditions ("skew low", "symmetric" and "skew high") and the two types of valuation tasks (risk aversion and individual discount rates).[13] We restrict the horizons in the discount rate task to three: 1, 4 and 6 months. The Switching MPL (sMPL) varies the standard MPL by asking the subject to choose which row he wants to switch at, assuming underlying monotonicity of the underlying preferences to fill out the remaining choices for the subject.[14] The sMPL is implemented because the iMPL changes the decision from the MPL in two ways: forcing a single switch point, and refining the choice. By comparing the MPL and sMPL we can see the pure effect of the first change, and by comparing the sMPL and the iMPL we can see the pure effect of the second change. Thus we have a 3x3x2 design. The two first treatments are implemented between-subjects, so that any one subject only experiences one of the three MPL institutions and one of the three frames. The last treatment is implemented within-subjects, such that each subject faces 4 risk aversion tasks with varying stakes and 3 discount rate tasks with varying horizons. The monetary incentives in the valuation tasks are the same as those applied in the field experiments.

In addition, we examined three treatments that are applied equally to all subjects in each session. One is a randomization of their initial endowment. Each subject received a guaranteed 250 DKK to participate, and we randomly assigned them an extra amount between 10 and 100 DKK, chosen from a discrete uniform distribution in increments of 10 DKK. The second treatment is a randomization of the order of presentation of each of the four risk aversion tasks and each of the three individual discount rate (IDR) tasks. The rationale for these treatments, and estimates of their effects, are presented in Anderson, Harrison, Lau and Rutström [2004].

---

[13] In the discount rate task, the *skewHI* treatment offers initial annual interest rates of (15%, 25%, 35%, 40%, 45%, and 50%), while the *skewLO* treatment offers annual interest rates of (5%, 10%, 15%, 25%, 35%, and 50%). The symmetric treatment offers 10 rows with annual interest rates between 5% and 50%.

[14] We believe that the first implementation of the enforced-single-switching feature of the sMPL was by Gonzales and Wu [1999].

3. Results

We examine the effect of our treatments on average measures of risk aversion and individual discount rates elicited in the field and lab experiments. Since we use the same valuation tasks in the field and lab experiments and collect the same information on socio-demographic variables, we can also investigate correlations between treatments and demographic characteristics in both samples. Detailed analyses of the full range of treatments in the field and lab experiments are reported in Harrison, Lau, Rutström and Sullivan [2005] and Andersen, Harrison, Lau and Rutström [2004], respectively.

We examine the "overlap" here, which are the skewness frame treatments in the risk aversion experiments and the three shorter time horizons in the discount rate experiments. In each case we retain controls for the additional treatments employed in the lab experiments (e.g., randomization of task order, use of random initial endowments).

Similarly, there are some demographic characteristics we simply did not observe in the lab but did observe in the field. For example, none of our lab subjects were retired, and none were older than 32. Again, we focus on the overlapping characteristics, and control for others when examining the field data.

Our hypotheses concern three aspects of the elicited value. First, are the average elicited values the same in the field and the lab? Second, are the estimates of demographic effects the same in the field and the lab? In other words, do we estimate the same effect of observable characteristics such as sex or age on the elicited value? And third, are the estimates of treatments effects the same in the field and the lab? Do we see the same effect of skewness frames on elicited risk attitudes, or the same effect of time horizon on elicited discount rates?

A. Measures of Risk Attitudes

Figure 1 shows the observed distribution of risk attitudes in our field and lab experiments, using the raw mid-point of final iterations of the elicited CRRA intervals. For comparability, the distributions only reflect the symmetric menu treatment. Using CRRA as the characterization of risk

attitudes, a value of 0 denotes risk neutrality, negative values indicate risk-loving, and positive values indicate risk aversion. Thus we see *comparable degrees of risk aversion in the field and lab*: the mean CRRA coefficient in the field sample is 0.63 and the median is 0.62, while the mean coefficient is 0.79 and the median is 0.80 in the lab sample.

Although it is comforting to see that *average* risk aversion estimates are comparable in the lab and the field, that does not address their *variation* with observable demographic characteristics and/or treatments. If there is significant variation in risk aversion with demographic characteristics then the equality of the averages is simply a coincidence of the demographic composition of the two samples. In order to assess the importance of demographics on risk attitudes we apply regression models that condition on observable characteristics of the subjects. Table 3 provides the definitions of the explanatory variables and summary statistics of the two samples. The lab sample is much younger, which is no surprise. Similarly, we see large differences in characteristics such as whether the subject has any children, whether they own their apartment or house, whether they are a student, the completed level of education, income levels, or resides in greater Copenhagen. But the lab sample also differs in other respects, which were unexpected: 51% of the field sample were female, compared to only 27% of the lab sample.

Results from the Field

Table 4 displays the results from estimating a panel interval regression model of the elicited CRRA values in the field experiments.[15] This model uses panel data since each subject provided four responses, one for each stake condition. Unobserved individual effects are modeled using a random-effects specification. The estimated model shown in Table 4 allows us to predict an estimate of the sample mean. For the symmetric frame this average is 0.68, which is close to the mean CRRA coefficient derived from the unconditional raw data shown in Figure 1.

Turning to the variation in risk aversion across our field sample, we consider the marginal

---

[15] All statistical analyses are undertaken using version 8.2 of *Stata*, documented in StataCorp [2003]. Our data and statistical commands are available at the *ExLab* Digital Library at http://exlab.bus.ucf.edu.

effect of individual demographics, holding constant the average value of other demographics. We find an effect on risk attitudes from age and education. 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 the age of 40 subjects become significantly less risk averse than those aged between 30 and 39, the omitted age group.[16] Skilled workers, which are those with some completed post-secondary education, have significantly higher aversion to risk than unskilled workers, and this higher aversion to risk is amplified if the subject has gone on to complete substantial higher education.

We also find that students generally exhibit significantly higher risk aversion than non-students, a result that is more notable since we separately control for pure age effects. Since the subjects were all 19 or older, they were all post-secondary students like most of the subjects attending our lab experiments. We thus observe that subjects with a completed post-secondary education, or a higher education, have higher aversion to risk than unskilled workers, and post-secondary students have higher risk aversion than non-students. Beginning or having completed post-secondary education is associated with higher aversion to risk than otherwise, which suggests that risk preferences may be in the same range if a subject's "state of nature" changes from being a student and unskilled to having completed a post-secondary education or higher.[17]

There is no significant effect from sex on risk attitudes. 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.[18]

---

[16] We can convincingly reject the hypothesis that all age effects are the same (p-value is 0.0312), as well as the null hypothesis that they are jointly zero (p-value is 0.0370).

[17] Using similar valuation tasks but smaller monetary incentives, Harrison, Johnson, McInnes and Rutström [2004] find that risk attitudes for the same individuals are stable over the six month period used in their lab experiments. To assess stability over longer periods of time requires that one take into account possible changes in the states of nature that individuals might condition their risk preferences over.

[18] 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. 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

The variables *skewLO* and *skewHI* control for the frame used. The *skewHI* treatment is statistically significant, but there is no significant effect from the *skewLO* treatment.[19] The *skewLO* treatment yields an average estimated CRRA of 0.43, and the *skewHI* treatment yields an average estimated CRRA of 0.95, each in a direction expected *a priori*. Both estimates are consistent with the conclusion that subjects in the field experiments are risk averse.

Results from the Lab

Table 5 reports results from a panel interval regression model of elicited CRRA values in our lab experiments, controlling for MPL institution, framing condition, task order, experimenter and individual demographics.[20] Unobserved individual effects are again modeled using a random-effects specification.

In general, we see a *substantially smaller effect of heterogeneity in our lab sample compared to the field sample*. The only (weakly) significant demographic variable is "single" status. The elicited CRRA value for singles is on average 0.24 lower than subjects with a partner, and this effect has a *p*-value of 0.06. We do not find a significant effect from age or education on elicited CRRA values in the lab sample, no doubt due to the homogeneity of the sample: 98% of the subjects are less than 30 years of age, 82% are students, and 78% have not completed some post-secondary education and are classified as unskilled workers.

Turning to estimates of treatment effects, the *framing of the decision table differs across the treatments in terms of significance*. The *skewLO* frame lowers average CRRA, by about 0.23, and this effect is statistically significant in the lab with a *p*-value of 0.046. However, in the field, this average effect was only to lower average CRRA by 0.027 and the *p*-value on that estimated effect was 0.803.

Grossman [2004] review these studies and several unpublished studies. Rubin and Paul [1979] and Robson [1996] offer evolutionary models of possible sex differences in risk aversion. Even if there is no evidence for an effect of sex on risk aversion, it is possible that observers may *predict* differences in risk attitudes based on sex (Eckel and Grossman [2002]).

[19] Harrison, Lau, Rutström and Sullivan [2005] 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 focuses 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.

[20] There are no observations on the two highest age groups and the retired group in our lab experiments, and three binary variables related to these demographics are therefore excluded in the model.

Similarly, the *skewHI* frame has no effect on average CRRA (the coefficient estimate is only 0.004 with a *p*-value of 0.972), compared to an increase of 0.27 in the field (*p*-value= 0.013). Thus subjects in the lab and the field respond differently to the framing treatments, but the effects are more or less consistent with the hypothesis that motivated the treatments.

Comparing the Field and the Lab

We find some significant differences between the elicited risk attitudes in the lab and the field. The lab results fail to detect the substantial preference heterogeneity that we find in the field, due to the demographic homogeneity of the lab. However, the lab results also lead us to draw strikingly different conclusions with respect to the statistical significance of the treatment effects. We do not claim that the "lab is wrong" and the 'field is right" with respect to the treatment effects, just that they are very different. It could, for example, be that the lab sample is more sensitive to treatment effects since we have mitigated sample variation by sampling from a more homogenous population (the "lab rats" argument). It is also possible that the lab sample differs from the field sample in additional ways that we are not able to observe.[21]

We can, however, investigate if there is an interaction effect between treatment effects and demographic effects. This will allow us to then say whether it is the lab or the field that is more likely to have the correct estimates of treatment effects, since the field sample includes variations in treatments *and* demographics. If we observe that there is some interaction between the two effects, and we have strong priors that the field is more representative in terms of demographic heterogeneity, then there will be a presumption that the lab is generating the wrong results.

This is, in fact, what we find. To illustrate, we only need consider two demographic characteristics: age and sex. We pick these two since we know that we *do* have considerable variation within both the lab and field samples in sex, and that we do *not* have much variation within the lab sample in age. To examine these interaction effects we include all interactions, re-estimate the

---

[21] We do find that there is significant variation in the lab sample responses that is not predicted by the estimated field model.

regression models reported in Tables 5 for the field sample, and then test for the absence of interaction effects formally.[22] In the field data we reject the hypothesis that females provided the same CRRA independently of frame, using a *p*-value of 0.058. The comparable test for males can only be rejected at a *p*-value of 0.17. Although borderline in terms of conventional significance levels, we would also reject the hypothesis that males and females provide the same CRRA within the symmetric frame at a *p*-value of 0.12. Thus there is evidence of interaction effects between sex and the treatments from the field data.

Turning to age, we find even more striking evidence of interaction effects. Within the *skewLO* frame, we reject the hypothesis that age does not matter with a *p*-value of 0.042; within the symmetric frame the same test generates a *p*-value of only 0.011. Particularly telling is the finding that these effects are driven by the older subjects in the field. We reject the hypothesis that middle-aged subjects, between 40 and 50, have the same CRRA when faced with different frames at a *p*-value of 0.0817. The same test using older subjects, aged over 50, is rejected at a *p*-value of 0.0004. Of course, since the lab sample contained none of these older subjects, it could not detect this interaction effect between treatments and age.

We conclude that there are significant interaction effects between the treatment effects and the demographic effects. Not only were these observed in the domain that our lab sample could never capture (age), but they were also observed in the domain that our lab sample could have captured (sex). These results clearly point to the value of going to the field in the manner that we did, *even if one is only interested in getting accurate estimates of treatment effects*. Since treatment effects may interact with individual characteristics, and since the field provides more variation in these characteristics, it also offers a more reliable testing ground.

---

[22] The detailed statistical results are reported in the output files available at the *ExLab* Digital Library at http://exlab.bus.ucf.edu, and are not of interest beyond the tests of interaction effects reported in the text. Each test is a χ^2 test that the appropriate coefficient estimates are identical.

B. Measures of Individual Discount Rates

Figure 2 displays the elicited discount rates for our subjects in the field and lab experiments, using the mid-point of the final interval selected. The discount rates are pooled across all horizons included in the lab experiments (1, 4 and 6 months), and these distributions only reflect the symmetric menu treatment in the lab.[23] The mean elicited discount rate in the field experiments is 25.0%, with a median of 26.1%.[24] The mean coefficient in the lab experiments is 27.9%, with a median of 26.8%. Thus we see *roughly comparable average measures of discount rates when we compare the lab and the field.*

Results from the Field

Table 6 displays the results from estimating a panel interval regression model of the elicited IDR values in the field experiments for the three horizons that are also included in the lab experiments. The default horizon is the 1-month task, so we observe that there is a small but statistically significant effect of horizon on elicited discount rates. Both 4-month and 6-month discount rates are about 3¾ percentage points lower than the 1-month average. The only demographic characteristic to have a statistically significant effect is living in the Copenhagen area, which increases average discount rates by 5.8 percentage points.[25] Higher education and low income have a weakly significant effect: the former decreasing average IDR by about 5 percentage points and the latter increasing it by about the same amount. Each is significant at roughly the 10% level. Although heterogeneity is an issue for discount rates, it does not appear to be as strong as it is for risk attitudes.

---

[23] We did not employ "skewing" variations of the frame in the field IDR experiments, so using the symmetric lab results is directly comparable to the field results.

[24] These values are virtually identical to those found in the field by HLRS: they estimated a mean of 23.1%, a median of 22.4%, and a standard deviation of 14.0%.

[25] The related variable capturing if someone lives in a larger city also increases average IDR, by 4.2 percentage points and with a *p*-value of 0.12.

<u>Results from the Lab</u>

Table 7 displays comparable results from the lab experiments. Treatment effects are roughly double those obtained in the field, and also statistically significant.[26] This suggests that *one* factor causing short-term discount rates to be higher than longer-term discount rates in lab settings is the simple fact that such behavior is more representative of the subjects one encounters in a lab setting than in a field setting. We do not claim that this explanation fully accounts for these differences, but that is it one factor that should be taken into account.[27]

Although no demographic effect is statistically significant at the 10% level, many have large coefficients. For example, those that have completed higher education have discount rates that are 8.1 percentage points lower than others, those with higher incomes have discount rates that are lower by 10.0 percentage points, and those living in Copenhagen by 11.0 percentage points.

<u>Comparing the Field and the Lab</u>

We therefore find similar but stronger treatment effects in the lab than in the field, but somewhat different demographic effects. The lab produces no significant demographic effects, including those found to be significant or borderline significant in the field (Copenhagen residence, longer education, and low income), although many characteristics appear to be correlated with large swings in average IDR in the lab.[28]

Turning to estimates of interaction effects, we again find that demographics affect the estimated responses to the treatments. The most striking result is that significant horizon differences

---

[26] Although not a focus of our analysis, it is interesting to note that there is a significant experimenter effect in the lab that was not found in the field data. In the lab it appears that experimenter Andersen was associated with elicited discount rates that were 7.5 percentage points lower than experimenter Lau, and this effect has a *p*-value of 0.064. This could be due to Lau being a little slower at conducting the experiments: over a session that generally required 2 hours, he typically required 15 to 20 minutes more. This could have encouraged some impatience in subjects that spilled over into their responses, although this is conjecture. The same differences in speed of executing the experiment were found in the field, but there we observe no significant effects from the experimenter (the average IDR is again lower with Andersen at the helm, but only by 2.9 percentage points and with a *p*-value of 0.19).

[27] See Coller, Harrison and Rutström [2003] for further discussion of the experimental issue involved in testing for constancy of discount rates with respect to time horizon.

[28] As is the case for risk attitudes, we find that there is significant variation in the lab sample responses to the inter-temporal decisions that are not predicted by the estimated field model.

are only observed for the two younger age groups in the field, those aged below 40. This interaction effect explains why the horizon effect appeared to be so much larger in the lab than in the field: the lab *only* contained those people that we observe having a significant horizon effect in the field, and *none* of the people that we observe having *no* horizon effect in the field. Thus the lab treatment effects, showing a much larger horizon effect, are an artefact of the skewed age distribution in the lab relative to the field. The lab results are not wrong, conditional on the age of the subjects in the lab, it is just that their ages provide a misleading indicator of how the treatments affect people of all ages.

4. Conclusions

Our results provide striking evidence that there are good reasons to conduct field experiments. Although we find that both average risk attitudes and discount rates are roughly the same in the lab and the field, this is primarily a coincidence based on the composition of the particular samples recruited. We find evidence of preference heterogeneity based on our field responses, which implies that a slight change in the composition of the lab sample could easily result in responses that on average differ from those elicited in the field. The homogeneity of the university student population limits the ability to detect the preference heterogeneity that is present in the field when using lab experiments. More importantly, there are differences in treatment effects measured in the lab and the field that can, at least in part, be traced to interactions between treatment and demographic effects. These can only be detected and controlled for properly in the field data. Thus one cannot simply claim, without additional empirical argument or assumption, that treatment effects estimated in the lab are reliable. The same would hold true for demographic effects estimated in lab experiments if there are important undiscovered interaction effects in the lab.

Our findings should not lead one to conclude that lab experiments in general are uninformative or misleading. The claim we make is simply that for behavior that is characterized by heterogeneity, such as subjective preferences, lab experiments can only provide limited information.

References

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Elicitation Using Multiple Price List Format," *Working Paper 04-8*, Department of Economics, College of Business Administration, University of Central Florida, 2004.

Blackburn, McKinley; Harrison, Glenn W., and Rutström, E. Elisabet, "Statistical Bias Functions and Informative Hypothetical Surveys," *American Journal of Agricultural Economics*, 76, December 1994, 1084-1088.

Botelho, Anabela V.; Harrison, Glenn W. ; Hirsch, Marc, and Rutström, E. Elisabet, "Bargaining Behavior, Demographics, and Nationality: A Reconsideration of the Experimental Evidence," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).

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.

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, May 2004.

Eckel, Catherine C., and Grossman, Philip J., "Forecasting Risk Attitudes: An Experimental Study of Actual and Forecast Risk Attitudes of Women and Men," *Unpublished Manuscript*, Department of Economics, Virginia Tech, October 2002.

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, 2004).

Gonzalez, Richard, and Wu, George, "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 1999, 129-166.

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.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Temporal Stability of Estimates of Risk Aversion," *Applied Financial Economics Letters*, 2004, forthcoming.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95, 2005 forthcoming.

Harrison, Glenn W.; Lau, Morten I; and Rutström, E. Elisabet, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Working Paper 04-07*, Department of Economics, College of Business Administration, University of Central Florida, 2004.

Harrison, Glenn W.; Lau, Morten I.; Rutström, E. Elisabet, and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).

Harrison, Glenn W.; Lau, Morten I., 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," *Journal of Economic Literature*, 42(4), December 2004, 1013-1059.

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.

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.

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.

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

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.

Saha, Atanu, "Expo-Power Utility: A Flexible Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75(4), November 1993, 905-913.

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.

StataCorp, *Stata Statistical Software: Release 8.0* (College Station, TX: Stata Corporation, 2003)

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.

Table 2: Payoff Table for 6 Month Time Horizon

Payoff Alternative	Payment Option A (pays amount below in 1 month)	Payment Option B (pays amount below in 7 months)	Annual Interest Rate (AR, in percent)	Annual Effective Interest Rate (AER, in percent)	Preferred Payment Option (Circle A or B)	
1	3,000 DKK	3,038 DKK	2.5	2.52	A	B
2	3,000 DKK	3,075 DKK	5	5.09	A	B
3	3,000 DKK	3,114 DKK	7.5	7.71	A	B
4	3,000 DKK	3,152 DKK	10	10.38	A	B
5	3,000 DKK	3,190 DKK	12.5	13.1	A	B
6	3,000 DKK	3,229 DKK	15	15.87	A	B
7	3,000 DKK	3,268 DKK	17.5	18.68	A	B
8	3,000 DKK	3,308 DKK	20	21.55	A	B
9	3,000 DKK	3,347 DKK	22.5	24.47	A	B
10	3,000 DKK	3,387 DKK	25	27.44	A	B
11	3,000 DKK	3,427 DKK	27.5	30.47	A	B
12	3,000 DKK	3,467 DKK	30	33.55	A	B
13	3,000 DKK	3,507 DKK	32.5	36.68	A	B
14	3,000 DKK	3,548 DKK	35	39.87	A	B
15	3,000 DKK	3,589 DKK	37.5	43.11	A	B
16	3,000 DKK	3,630 DKK	40	46.41	A	B
17	3,000 DKK	3,671 DKK	42.5	49.77	A	B
18	3,000 DKK	3,713 DKK	45	53.18	A	B
19	3,000 DKK	3,755 DKK	47.5	56.65	A	B
20	3,000 DKK	3,797 DKK	50	60.18	A	B

Table 3: List of Variables and Descriptive Statistics

Variable	Definition	Raw Field Sample Mean	Raw Lab Sample Mean
female	Female	0.51	0.27
young	Aged less than 30	0.17	0.98
middle	Aged between 40 and 50	0.28	0.00
old	Aged over 50	0.37	0.00
single	Lives alone	0.20	0.48
kids	Has children	0.28	0.02
nhhd	Number of people in the household	2.49	1.98
owner	Owns own home or apartment	0.69	0.19
retired	Retired	0.16	0.00
student	Student	0.09	0.82
skilled	Some post-secondary education	0.38	0.11
longedu	Substantial higher education	0.36	0.14
IncLow	Lower level income	0.34	0.74
IncHigh	Higher level income	0.33	0.16
copen	Lives in greater Copenhagen area	0.27	0.87
city	Lives in larger city of 20,000 or more	0.39	0.09
experimenter	Experimenter Andersen (default is Lau)	0.49	0.66

Legend: Most variables have self-evident definitions. The omitted age group is 30-39. Variable "skilled" indicates if the subject has completed vocational education and training or "short-cycle" higher education, and variable "longedu" indicates the completion of "medium-cycle" higher education or "long-cycle" higher education. These terms for the cycle of education are commonly used by Danes (most short-cycle higher education program last for less than 2 years; medium-cycle higher education lasts 3 to 4 years, and includes training for occupations such as a journalist, primary and lower secondary school teacher, nursery and kindergarten teacher, and ordinary nurse; long-cycle higher education typically lasts 5 years and is offered at Denmark's five ordinary universities, at the business schools and various other institutions such as the Technical University of Denmark, the schools of the Royal Danish Academy of Fine Arts, the Academies of Music, the Schools of Architecture and the Royal Danish School of Pharmacy). Lower incomes are defined in variable "IncLow" by a household income in 2002 below 300,000 kroner. Higher incomes are defined in variable "IncHigh" by a household income of 500,000 kroner or more.

Table 4: Statistical Model of Risk Aversion Responses in Field Experiments

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

N=925, based on 245 subjects.

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		0.26	0.30	0.37	-0.32	0.84
skewLO	SkewLO frame	-0.03	0.11	0.80	-0.24	0.19
skewHI	SkewHI frame	0.27	0.11	0.01	0.06	0.49
PrizeSet2	Second prize set	0.28	0.06	0.00	0.17	0.39
PrizeSet3	Third prize set	0.18	0.06	0.00	0.07	0.29
PrizeSet4	Fourth prize set	0.19	0.06	0.00	0.08	0.30
experimenter	Experimenter effect	-0.06	0.09	0.53	-0.24	0.12
female	Female	-0.08	0.09	0.38	-0.26	0.10
young	Aged less than 30	0.15	0.18	0.42	-0.20	0.49
middle	Aged between 40 and 50	-0.29	0.14	0.04	-0.56	-0.01
old	Aged over 50	-0.10	0.16	0.52	-0.43	0.22
single	Lives alone	0.02	0.15	0.92	-0.28	0.31
kids	Has children	0.05	0.14	0.74	-0.23	0.32
nhhd	Number in household	-0.01	0.06	0.94	-0.13	0.12
owner	Owns home or apartment	0.05	0.13	0.72	-0.20	0.29
retired	Retired	-0.10	0.15	0.49	-0.40	0.19
student	Student	0.33	0.18	0.07	-0.02	0.68
skilled	Some post-secondary education	0.28	0.12	0.02	0.04	0.52
longedu	Substantial higher education	0.34	0.13	0.01	0.09	0.59
IncLow	Lower level income	-0.02	0.13	0.86	-0.27	0.23
IncHigh	Higher level income	0.03	0.12	0.83	-0.21	0.26
copen	Lives in Copenhagen area	0.20	0.12	0.10	-0.04	0.45
city	Lives in larger city of 20,000 or more	0.18	0.11	0.11	-0.04	0.40
σ_u	Standard deviation of individual effect	0.61	0.04	0.00	0.54	0.68
σ_e	Standard deviation of residual	0.57	0.02	0.00	0.54	0.61

Notes: Log-likelihood value is -2532.4; Wald test for null hypothesis that all coefficients are zero has a χ^2 value of 74.15 with 22 degrees of freedom, implying a p-value of less than 0.0001; fraction of the total error variance due to random individual effects is estimated to be 0.53, with a standard error of 0.034.

Table 5: Statistical Model of Risk Aversion Responses in Lab Experiments

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

N=354, based on 90 subjects.

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		0.77	0.38	0.04	0.03	1.51
smpl	sMPL format	0.02	0.11	0.88	-0.19	0.23
impl	iMPL format	0.37	0.11	0.00	0.15	0.58
skewLO	SkewLO frame	-0.23	0.12	0.05	-0.46	-0.00
skewHI	SkewHI frame	-0.00	0.11	0.97	-0.22	0.21
Task2	Second task	0.05	0.05	0.33	-0.05	0.14
Task3	Third task	0.03	0.05	0.55	-0.07	0.12
Task4	Fourth task	0.14	0.05	0.00	0.05	0.24
PrizeSet2	Second prize set	0.11	0.05	0.02	0.02	0.21
PrizeSet3	Third prize set	0.07	0.05	0.15	-0.02	0.17
PrizeSet4	Fourth prize set	0.04	0.05	0.42	-0.05	0.13
experimenter	Experimenter effect	-0.09	0.10	0.36	-0.27	0.10
endowment	Initial endowment	-0.00	0.00	0.32	-0.00	0.00
female	Female	0.09	0.10	0.38	-0.11	0.30
single	Lives alone	-0.24	0.13	0.06	-0.49	0.01
nhhd	Number in household	0.00	0.08	0.99	-0.15	0.15
owner	Owns home or apartment	0.11	0.15	0.44	-0.18	0.41
student	Student	-0.04	0.12	0.73	-0.27	0.19
skilled	Some post-secondary education	-0.03	0.15	0.84	-0.32	0.26
longedu	Substantial higher education	-0.03	0.13	0.81	-0.28	0.22
IncLow	Lower level income	0.05	0.17	0.76	-0.29	0.39
IncHigh	Higher level income	-0.07	0.22	0.74	-0.50	0.35
copen	Lives in Copenhagen area	0.07	0.23	0.75	-0.38	0.53
city	Lives in larger city of 20,000 or more	0.00	0.28	0.99	-0.54	0.55
σ_u	Standard deviation of individual effect	0.36	0.03	0.00	0.30	0.43
σ_e	Standard error of residual	0.28	0.01	0.00	0.25	0.31

Notes: Log-likelihood value is -558.4; Wald test for null hypothesis that all coefficients are zero has a χ^2 value of 50.47 with 23 degrees of freedom, implying a p-value of 0.0008; fraction of the total error variance due to random individual effects is estimated to be 0.63, with a standard error of 0.05.

Table 6: Statistical Model of IDR Responses in Field Experiments

Random-effects interval regression,
with the final discount rate interval chosen by the subject as the dependent variable.

N=756, based on 252 subjects.

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		32.39	7.11	0.00	18.45	46.33
horizon4	4 months horizon	-3.64	1.10	0.00	-5.79	-1.49
horizon6	6 months horizon	-3.90	1.10	0.00	-6.06	-1.75
experimenter	Experimenter effect	-2.95	2.28	0.19	-7.41	1.51
female	Female	-0.02	2.25	0.99	-4.43	4.40
young	Aged less than 30	-4.95	4.45	0.27	-13.67	3.77
middle	Aged between 40 and 50	2.05	3.47	0.55	-4.75	8.84
old	Aged over 50	3.25	4.06	0.42	-4.70	11.21
single	Lives alone	-0.95	3.62	0.79	-8.05	6.15
kids	Has children	4.38	3.52	0.21	-2.52	11.27
nhhd	Number in household	-0.91	1.56	0.56	-3.96	2.15
owner	Owns home or apartment	0.11	3.05	0.97	-5.87	6.08
retired	Retired	-4.26	3.62	0.24	-11.35	2.83
student	Student	-1.18	4.44	0.79	-9.88	7.53
skilled	Some post-secondary education	-2.04	3.01	0.50	-7.93	3.85
longedu	Substantial higher education	-5.16	3.11	0.10	-11.25	0.93
IncLow	Lower level income	5.07	3.10	0.10	-1.00	11.14
IncHigh	Higher level income	-1.34	2.91	0.65	-7.04	4.37
copen	Lives in Copenhagen area	5.77	3.04	0.06	-0.19	11.73
city	Lives in larger city of 20,000 or more	4.22	2.74	0.12	-1.14	9.58
σ_u	Standard deviation of individual effect	15.39	0.88	0.00	13.66	17.12
σ_e	Standard deviation of residual	11.81	0.40	0.00	11.02	12.60

Notes: Log-likelihood value is -3392.7; Wald test for null hypothesis that all coefficients are zero has a χ^2 value of 32.85 with 19 degrees of freedom, implying a p-value of 0.025; fraction of the total error variance due to random individual effects is estimated to be 0.63, with a standard error of 0.032.

Table 7: Statistical Model of IDR Responses in Lab Experiments

Random-effects interval regression,
with the final discount rate interval chosen by the subject as the dependent variable.

N=270, based on 90 subjects.

Variable	Description	Estimate	Standard Error	p-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		51.55	16.11	0.00	19.97	83.14
horizon4	4 months horizon	-6.74	1.67	0.00	-10.01	-3.48
horizon6	6 months horizon	-9.07	1.66	0.00	-12.32	-5.82
smpl	sMPL format	-3.85	4.52	0.39	-12.71	5.01
impl	iMPL format	1.85	4.64	0.69	-7.24	10.95
Task2	Second task	3.51	1.66	0.03	0.25	6.77
Task3	Third task	2.17	1.65	0.19	-1.06	5.40
skewLO	SkewLO frame	3.54	4.86	0.47	-5.98	13.06
skewHI	SkewHI frame	3.85	4.59	0.40	-5.14	12.84
experimenter	Experimenter effect	-7.37	4.03	0.07	-15.27	0.54
endowment	Initial endowment	-0.01	0.07	0.90	-0.14	0.12
female	Female	2.45	4.42	0.58	-6.21	11.12
single	Lives alone	-2.50	5.35	0.64	-12.98	7.99
nhhd	Number in household	3.52	3.21	0.27	-2.78	9.82
owner	Owns home or apartment	-7.83	6.25	0.21	-20.08	4.42
student	Student	-6.30	4.98	0.21	-16.06	3.46
skilled	Some post-secondary education	-6.86	6.19	0.27	-18.98	5.27
longedu	Substantial higher education	-8.09	5.38	0.13	-18.64	2.47
IncLow	Lower level income	-3.57	7.38	0.63	-18.04	10.90
IncHigh	Higher level income	-10.04	9.11	0.27	-27.90	7.83
copen	Lives in Copenhagen area	-11.02	9.97	0.27	-30.56	8.52
city	Lives in larger city of 20,000 or more	-1.80	11.81	0.88	-24.94	21.35
σ_u	Standard deviation of individual effect	15.58	1.43	0.00	12.79	18.38
σ_e	Standard deviation of residual	10.43	0.62	0.00	9.22	11.64

Notes: Log-likelihood value is -708.7; Wald test for null hypothesis that all coefficients are zero has a χ^2 value of 58.26 with 21 degrees of freedom, implying a p-value of less than 0.0001; fraction of the total error variance due to random individual effects is estimated to be 0.69, with a standard error of 0.048.

Figure 1: Distribution of CRRA in Denmark
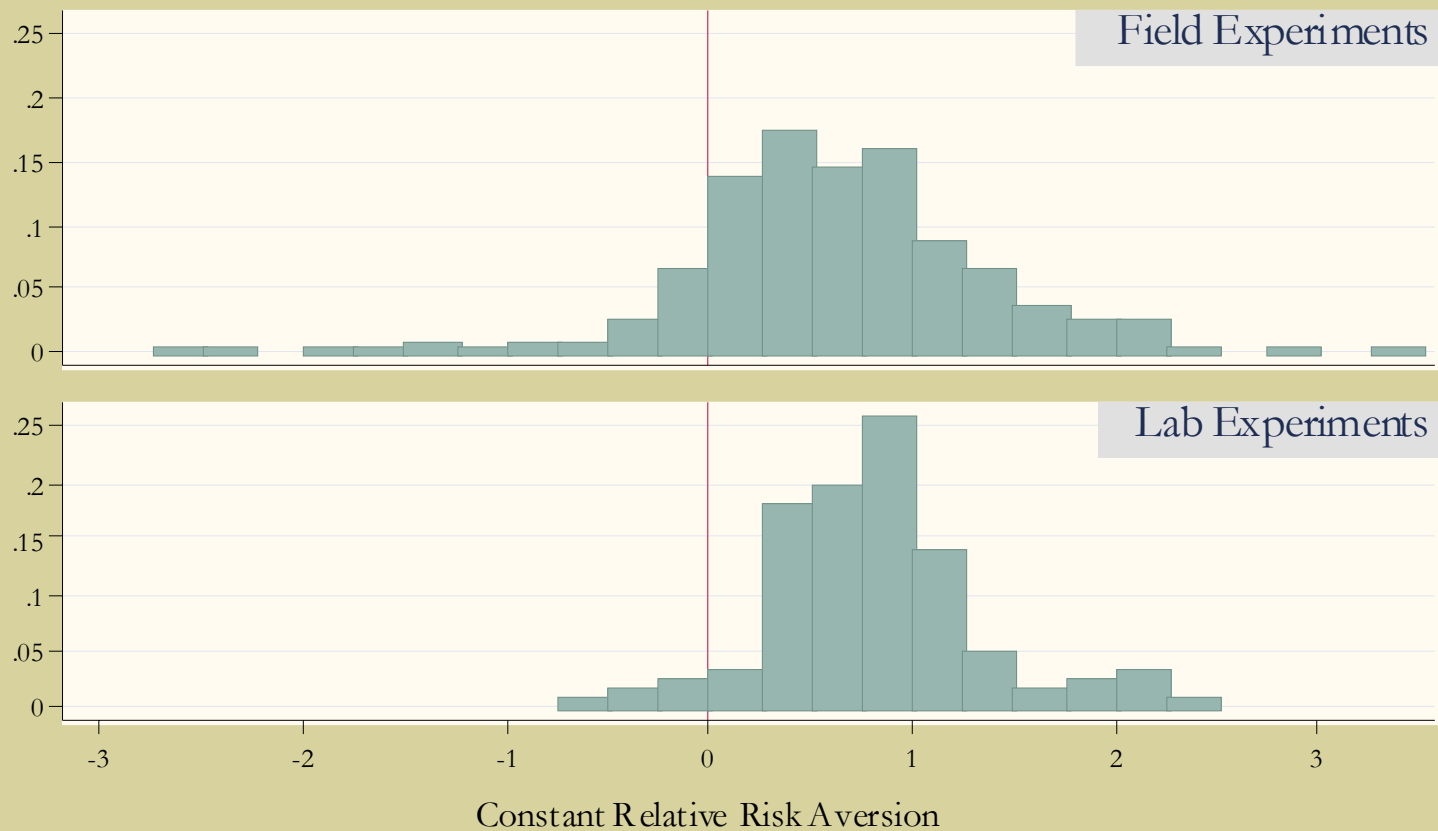Mid-Point of Raw Response Interval with Symmetric Menu

Figure 2: Distribution of IDR in Denmark
Mid-Point of Raw Response Interval