Measurement With Experimental Controls

by

Glenn W. Harrison, Eric Johnson, Melayne M. McInnes & E. Elisabet Rutström†

October 2006

Forthcoming in M. Boumans (ed.),
Measurement in Economics: A Handbook (San Diego, CA: Elsevier, 2007)

Abstract. Many predictions of economic theory depend on the assumed aversion of individuals towards risk. We examine statistical aspects of controlling for risk aversion in the lab, and the implications that these have on the ability to test expected utility theory. The concerns expressed here regarding the importance and difficulty of generating precise estimates of individual risk attitudes generalize to a wide range of other individual characteristics, such as inequality aversion and trust. We show that imprecision in estimated individual characteristics may result in misleading conclusions in tests of the underlying theory of choice. We also show that the popular instruments and statistical models used to estimate risk attitudes do not allow sufficiently precise estimates. Given existing laboratory technology and statistical models, we conclude that controls for risk aversion should be implemented using within-subjects, "revealed preference" designs that utilize the direct, raw responses of the subject. These statistical issues are generally applicable to a wide variety of experimental situations.

† Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström), Department of Economics, Kent State University, USA (Johnson), and Department of Economics, Moore School of Business, University of South Carolina, USA (McInnes). E-mail: GHARRISON@RESEARCH.BUS.UCF.EDU, EJOHNSON@BSA3.KENT.EDU, MCINNES@MOORE.SC.EDU and ERUTSTROM@BUS.UCF.EDU. We thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518 and NSF/POWRE 9973669 to Rutström, grant NSF/SES 0213974 to McInnes, and grants NSF/DRU 0527675 and NSF/SES 0616746 to Harrison and Rutström. We are grateful to Maribeth Coller, Morten Lau, Graham Loomes, Chris Starmer, Robert Sugden, Melonie Sullivan and Peter Wakker for comments, as well as participants at the 11th Conference on the Foundations & Applications of Utility, Risk & Decision Theory in Paris. Supporting data and instructions are stored at the ExLab Digital Archive at http://exlab.bus.ucf.edu.

Experimental methods provide the promise that economists will be able to measure latent concepts with greater reliability. The reason is that experimental methods offer the possibility of controlling potential confounds.

However, the use of experimental controls might not lead to more reliable measurements. One reason is that the imposition of an artefactual control might itself lead to changes in behavior compared to the naturally occurring counterpart of interest. Concern with this problem has spurred interest in field experiments, where the controls are less artificial than in many laboratory experiments.[1] It has also spurred renewed interest in sample selection and sorting processes.[2]

Another reason that experimental controls might not generate more reliable measurements is that the latent data-generating process might simply be mis-specified. If the experimental design is motivated by a model of the data-generating process that is invalid, then there can be no expectation that the controls will improve measurement and inference. For example, if there are actually two or more distinct data-generating processes at work, and we assume one, then systematically invalid inferences can result.[3]

We consider a third way in which experimental controls might influence measurement inference, by allowing "unobservables" to become "observable." Concepts that previously needed to be assumed to take on certain values or distributions *a priori*, can now be measured and controlled. In turn, this allows conditional measurements to be made unconditionally, akin to the integration of "nuisance parameters" in Bayesian analysis. We consider a substantively fundamental application of these ideas, to the evaluation of choice behavior under uncertainty when one has experimental control of risk attitudes.[4]

---

[1] Harrison and List [2004] review this literature, and this concern with laboratory experiments.

[2] For example, see Botelho, Harrison, Pinto and Rutström [2005], Harrison, Lau and Rutström [2005], Kocher, Strauß and Sutter [2006] and Lazear, Malmendier and Weber [2006].

[3] Harrison and Rutström [2005] illustrate this point by comparing estimates of choice behavior when either expected utility theory or prospect theory are assumed to generate the observed data, and contrast the results with those obtained from a finite mixture model that allows both to be valid for distinct sub-samples. Similarly, Coller, Harrison and Rutström [2006] compare inferences about temporal discounting models when one assumes that subjects discount exponentially or quasi-hyperbolically, when the data is better characterized by again assuming that distinct sub-sample follow each model.

[4] Other applications include Harrison [1990], Engle-Warnick [2004] and Karlan and Zinman [2005].

Many predictions of economic theory depend on the assumed aversion of individuals towards risk. Empirical research requires that one make a maintained assumption about risk attitudes or devise controls for risk aversion. The first strategy has the obvious disadvantage that the maintained assumption may be false. The second strategy is becoming feasible, particularly with the development of simple pre-tests for risk aversion in laboratory settings. We examine statistical aspects of controlling for risk aversion in the lab, and the implications that these have on the ability to test expected utility theory (EUT). The concerns expressed here regarding the importance and difficulty of generating precise estimates of individual risk attitudes generalize to a wide range of other individual characteristics, such as inequality aversion and trust.[5] Imprecision in estimated individual characteristics may result in misleading conclusions in tests of any underlying theory of choice.

The way in which controls for risk aversion can be implemented varies with the experimental design. If a "within-subjects" design is used, in which the same subject takes part in a risk aversion test and some other task, one can directly use the results for that subject to control for theoretical predictions in the other task. In general such responses are likely to respect the individual heterogeneity that one would expect from risk aversion, which is after all a subjective preference. If a "between-subjects" design is used, in which different subjects[6] are sampled for the risk aversion test and the other task, one must construct a statistical instrument for the risk attitudes of subjects in the latter task.

Instruments for risk attitudes can be generated by constructing a statistical model of risk attitudes from the responses to the risk aversion task, and then using that model to predict the attitudes of the subjects in the other tasks. Given the time and monetary cost of eliciting risk attitudes in addition to some other experimental task, such methodological short-cuts would be

[5] Methodologically related experimental procedures are being used to identify the extent of "inequality aversion" in tests of the propensity of individuals to "trust" each other. Cox [2004] discusses the need for controls in experiments in this context.

[6] We will assume that these are two samples drawn from the same population, and that there are no sample selection biases to worry about. These potential complications are not minor.

attractive to experimenters if reliable. Of course, relying on a statistical model means that one must recognize that there is some sampling error surrounding the estimated risk attitude, even if one assumes that the correct specification has been used for the statistical model.

The issue of imprecision in measuring risk is readily apparent when one uses a statistical model to predict risk parameters. While less obvious, this issue still arises when one uses a "direct test" in a within subjects design. Imprecision in directly eliciting risk aversion may arise for several reasons.[7] First, our risk elicitation task may not yield precise estimates due to "trembling hand" error on behalf of the subject. For example, even when given a simple choice between two lotteries, a subject may, with some positive probability, indicate one lottery when they intended to choose the other. A second source of error occurs if our risk attitude task does not elicit enough information to make sufficiently precise inferences about the parameters of the choice model. We can reduce the imprecision by improving the design of the risk elicitation task, but we still need a way to characterize the degree of imprecision in the estimated parameters and to gauge its impact on any conclusions that can be drawn based on responses in the subsequent choice task.[8]

We illustrate these procedures using a test of EUT as the "choice task" for which one needs a control for risk aversion.[9] We use data from a previous experiment in which subject choices have been shown to be inconsistent with expected utility given the estimate of the individual's risk

[7] "Imprecision" here is used in the usual econometric sense of minimizing the confidence intervals for the underlying parameter.

[8] Moffatt [2007] uses the concept of D-Optimal design to maximize the overall information content of an experiment.

[9] Rabin [2000] examines the theoretical role of risk aversion and EUT, and 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. He proves a calibration theorem 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. Following the interpretation of these arguments by Cox and Sadiraj [2006] and Rubinstein [2002], 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. If, on the other hand, one assumes utility is defined over income, this critique does not apply. A close reading of Rabin [2000; p. 1288] is consistent with this perspective, as is the model proposed by Charness and Rabin [2002] to account for experimental data they collect. 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 its 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. Fudenberg and Levine [2006] provide another framework for reconciling the EUTi and EUTw approaches, by positing a "dual self" model of decision-making in which a latent EUTw-consistent self constrains choices actually observed by the EUTi-motivated self.

attitude. We begin our analysis by allowing for the possibility that subjects are noisy decision-makers. One way to incorporate subject errors is to calculate their cost and ignore those inconsistent choices that have a trivial error cost. We show that the percentage of choices violating EUT remains high even when we consider only those errors that are costly to the subject. We then ask whether these results are sensitive to the precision with which we estimated an individual subject's risk attitude. The data we use were implemented using a full within-subjects design, allowing us to compare the use of direct, raw risk aversion measures for each subject in the EUT task with the use of instruments generated by a statistical model. The method we adopt is to examine the sensitivity of our conclusions about EUT to small perturbations in the estimated risk preference parameters. We can think of this test of *empirical sensitivity* as a counterpart to the formal sensitivity tests proposed by Magnus [2007]. Leamer [1978; p.207] and Mayer [2007] remind us to consider both economic significance as well as statistical significance when evaluating estimates of a parameter of interest. Our objective is to evaluate whether our economic conclusions are sensitive to small changes in the estimated parameters. In the case of our statistical model, the estimated confidence interval provides the standard region in which to conduct the perturbation study. When we use direct measures of risk, the nature of the experiment suggests natural regions in which to check for parameter sensitivity. These methods can also be used to address the issue of precision in tests of choice models other than EUT. Tests of cumulative prospect theory, for example, are conditional on the estimated parameters of the choice function and the robustness of the conclusions will depend on the precision with which the initial parameters were estimated.[10]

In Section 1 we review the need for estimates of risk attitudes in tests of EUT and show how inference is affected by risk aversion. In Section 2 we discuss experimental procedures for characterizing risk attitudes due to Holt and Laury [2002]. We present the distribution of estimated risk attitudes of our subject pool and examine its implications for subjects' preferences over lottery

[10] For example Harbaugh, Krause, and Vesterlund [2002] test the fourfold patten of risk attitude predicted by cumulative prospect theory. Their tests are designed based on parameter estimates from Prelec [1998] and others.

choices taken from the EUT choice experiments. In Section 3 we show that, for each subject and each choice, the cost of choosing inconsistently with EUT can be calculated. Conditional on our risk aversion estimates, we find that subjects frequently violate EUT even when the cost of doing so is high. In Section 4 we examine whether our risk aversion estimates provide sufficient precision for reaching meaningful conclusions about EUT. We show that the use of instruments, based on a statistical model, does not allow sufficiently precise estimates of risk aversion for our purposes. We discuss various reasons for this outcome. The implication is, however, that with existing laboratory technology and statistical models, controls for risk aversion should be implemented using within-subjects designs that utilize the direct, raw responses of the subject.[11]

1. Risk Aversion and Tests of EUT

Experiments that test EUT at the level of the individual typically require that the subject make two choices, so that we can compare their consistency. The first lottery choice can be used to infer the subject's risk attitude, and then the second choice can be used to test EUT, conditional on the risk attitude of the subject. Thus, preferences have to be elicited over *two* pairs of lotteries for there to be a test of EUT at all.

For a specific example of the frequently used "Common Ratio" (CR) test, suppose Lottery A consists of prizes $0 and $30 with probabilities 0.2 and 0.8. Lottery B, consisting of prizes $0 and $20 with probabilities 0 and 1. Then one may construct two additional compound lotteries, A* and B*, by adding a front end probability $q = 0.25$ of winning zero to lotteries A and B. That is, A* offers a $(1-q)$ chance to play lottery A and a q chance of winning zero. Subjects choosing A over B and B* over A*, or choosing B over A and A* over B*, are said to violate EUT.

[11] To what extent do our conclusions transfer beyond the lab to field experiments? In that setting one often encounters apologies that it was not possible to control everything of theoretical interest, but that the tradeoff was worth it because one is able to make more "externally valid" inferences about behavior. Such claims should be viewed with suspicion, and are often made just to hide incomplete experimental designs (Harrison [2005]). One can always condition on *a priori* distributions that might have been generated by other samples in more controlled settings (e.g., Harrison [1990]), such that inferences based on posteriors do not ignore that conditioning information. Or one can complement "uncontrolled" field experiments with controlled lab experiments, acknowledging that controls for risk attitudes in the latter might interact with other differences between the lab and the field.

To show precisely how risk aversion does matter, assume that risk attitudes can be characterized by the popular Constant Relative Risk Aversion (CRRA) function, $U(m) = (m^{1-r})/(1-r)$, where r is the CRRA coefficient. The certainty equivalents (CE) of the lottery pairs AB and A*B* as a function of r are shown in the left and right upper panels respectively of Figure 1. The CRRA coefficient ranges from -0.5 (moderately risk loving) up to 1.25 (very risk averse), with a risk-neutral subject at $r = 0$. The CE of lottery B, which offers $20 for sure, is the horizontal line in the left panel. The CE of A, A* and B* all decline as risk aversion increases. The lower panels of Figure 1 show the CE differences between the A and B (A* and B*) lotteries. Note that for the AB (A*B*) lotteries, the preferred outcome switches to lottery B (B*) for a CRRA coefficient about 0.45.

Most evaluations of EUT acknowledge that one cannot expect any theory to predict perfectly, since any violation would lead one to reject the theory no matter how many correct predictions it makes. One way to evaluate mistakes is to calculate their costs under the theory being tested and to "forgive" those mistakes that are not very costly while holding to account those that are. For each subject in our data and each lottery choice pair, we can calculate the CE *difference* given the individual's estimated CRRA coefficient allowing us to identify those choice pairs that are most salient. A natural metric for defining "trivial EUT violations" can then be defined in terms of choices that involve a difference in CE below some given threshold.

Suppose for the moment that an expected utility maximizing individual will flip a coin to make a choice whenever the difference in CE falls below some cognitive threshold. If $r = 0.8$, the CE difference in favor of B is large in the first lottery pair and B will be chosen. In the second lottery pair, the difference between the payoffs for choosing A* and B* is trivial[12] and a coin is flipped to make a choice. Thus, with probability 0.5 the experimenter will observe the individual choosing B and A*, a choice pattern inconsistent with EUT. In a sample with these risk attitudes, half the choices observed would be expected to be inconsistent with EUT.[13] With such a large

---

[12] In fact, it is less than 1 cent.

[13] This example illustrates only one possible characterization of subject error. For a discussion of alternatives, see Loomes, Moffatt, and Sugden [2002], Harless and Camerer [1994], Hey and Orme [1994], Hey [1995], Loomes and

difference between the choice frequencies, standard statistical tests would easily reject the hypothesis that they are the same. Thus, we would reject EUT in this case *even though EUT is true*.

Harrison, Johnson, McInnes and Rutström (HJMR) [2003] test EUT conditional on subjects' estimated risk aversion. Because the effects of risk aversion on our ability to test EUT depend on the particular lottery pairs used, HJMR [2003] consider 6 lottery pair choices and 12 stated "selling prices" for each of those 12 lotteries taken from Grether and Plott [1979], in addition to the two CR lotteries discussed above. In this literature on Preference Reversals (PR), the term "P-bets" refers to lotteries that have a relatively high *probability* of winning a relatively low monetary prize. The alternative bets are called "$-bets" because they have a lower probability of winning than the paired P-bet, but a higher *dollar* prize if the subject wins. PR Choice pair #1, for example, consists of the P-bet offering $4 with probability 35/36 and a loss of $1 with probability 1/36, and the $-bet offering $16 with probability 11/36 and a loss of $1 with probability 25/36. In this case the expected values are $3.86 and $3.85, respectively. The other five PR lottery pairs are similar. Because the lottery pairs in these experiments have virtually identical expected values, the difference in CE is zero if the CRRA coefficient $r = 0$.[14] Figure 2 shows that the difference in CE varies over the 6 lotteries.[15]

Based on the observed distribution of risk attitudes in our sample, we can calculate the EUT consistent choice in each of the 8 lottery pairs, assuming a CRRA utility function.[16] Abstracting from any consideration of the size of the CE difference, only 52% of observed choices were consistent with EUT when pooling over all 8 lottery pairs. Only in one PR lottery pair does one see a proportion of choices that are a markedly higher than 50% and therefore consistent with EUT. However, even those observations would require us to have a high tolerance for errors in the data in

---

Sugden [1995][1998], Ballinger and Wilcox [1997], and Carbone [1997].

[14] The payoffs in choice pair #6 have been altered slightly from the values in Grether and Plott [1979] so that a risk neutral individual is not indifferent between the two.

[15] Moffat [2007] develops a technique to select the parameters of the experiment to maximize the information content. Rather than use his technique, we adopted familiar tests from the literature, which has the advantage of allowing comparisons to previous findings.

[16] There are 8 lotteries in total, 6 PR lotteries and 2 CR lotteries. Each individual was presented with all 6 PR choices but only one of the CR choices. Hence, we have 7 observed choices for each individual. The analysis in HJMR [2003] was undertaken in the usual manner from the "preference reversal" literature: the direct binary choices of each subject were compared to the implied choices from the selling prices elicited over the underlying lotteries.

order to accept EUT. One would therefore reject the predictions of EUT for this set of choices, *conditional on the point estimates of risk aversion being accepted.*

While we observe a high rate of lottery choices inconsistent with EUT, this analysis does not consider whether the apparent errors are costly from the perspective of the subject. We ask here whether consistency with EUT increases as the cost of an error increases. Moreover, we investigate whether our rejection of EUT, conditional on the point estimates of risk aversion being accepted. is sensitive to the precision of our estimates of the underlying CRRA coefficients. Since the calculation of the CE differences depends critically on the CRRA estimates, we need to measure the robustness of our EUT findings to the imprecision of those estimates. While we focus on tests of EUT, this question of precision in estimating the parameters of the choice function is also relevant for tests of cumulative prospect theory, models of choice with altruism, other regarding preferences, etc.

One might ask how one can *test* EUT when one must *assume* that EUT holds in order to measure risk attitudes. Our point is that tests of EUT are incomplete if they do not also include a *joint hypothesis* about risk attitudes and consistent behavior over lottery choices. That is, one has to undertake such tests jointly or else one cannot test EUT at all, since the subject might be indifferent to the choices posed. Or, more accurately, without such tests of risk attitudes, the experimenter is unable to claim that he knows that the subject is not indifferent. So, just as EUT typically entails consistent behavior across two or more pairs of lottery comparisons, we are arguing it necessarily entails consistent behavior across those lottery comparisons *and* a task to measure risk attitudes. Our focus, then, is on the statistical precision of the inferences from the latter task.

2. Experimental Procedures

We use data from experiments reported in Harrison, Johnson, McInnes and Rutstrom [2003]. These experiments implement both a risk elicitation task and several lottery choices following those used in earlier experimental tests of the CR and PR phenomena.

The risk elicitation task follows Holt and Laury [2002] (HL) who devise a simple experimental measure for risk aversion using a multiple price list design. 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. 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 panel as $1.64, although the EV columns were not presented to subjects.[17] 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-to-last row. Arguably, the last row is simply a test that the subject understood the 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.

HL examine two main treatments designed to measure the effect of varying incentives.[18] They vary the scale of the payoffs in the matrix shown in Table 1 by multiplying the payoffs by 20, 50, or 90. Thus, Table 1 shows the scale of 1.

HJMR [2003] adapt the HL procedure by scaling it appropriately for the present purposes. Multiplying by 10 the original payoff scale of 1, which has prizes ranging between $0.10 and $3.85,

[17] There is an interesting question as to whether they should be provided. Arguably the subjects are trying to calculate them anyway, so providing them avoids a test of the joint hypothesis that "the subjects can calculate EV in their heads and will not accept a fair actuarial bet." On the other hand, providing them may cue the subjects to adopt risk-neutral choices. The effect of providing EV information deserves empirical study.

[18] HL's design provides in-sample tests of the hypothesis that risk aversion does not vary with income, an important issue for those that assume specific functional forms such as CRRA or constant absolute risk aversion (CARA), where the "constant" part in CRRA or CARA refers to the scale of the choices. A rejection of the "constancy" assumption is not a rejection of EUT in general, of course, but just these particular (popular) parameterizations.

provides responses that span prizes between \$1.00 and \$38.50. These two payoffs scales are referred to as 1x and 10x hereafter. The 10x payoffs comfortably covers the range of prizes needed to apply the measures of risk aversion to our experiments. All subjects were given the 10x test, but some were also given a 1x test prior to the 10x, which we refer to as the 1x10x treatment since these payoffs are comparable to the EUT decision tasks.[19]

Apart from conducting experiments to elicit subjects attitudes toward risk (the risk aversion experiments) HJMR [2003] also conducted experiments with the same subjects in order to test for violations of EUT, controlling for risk aversion. To avoid possible intra-session effects, only one experiment was run in each session. The same subjects were contacted again by e-mail and invited to participate in subsequent experiments that were separated by at least one week.[20] Students were recruited from the University of South Carolina. In total, 152 subjects participated in a risk aversion experiment and, of those, 88 also participated in the lottery choice experiments. Overall, there were 88 subjects for whom we can match results from the risk aversion test to the lottery choice task. No attempt was made to screen subjects for recruitment into subsequent experiments based on their choices in earlier experiments.

All subjects received a fixed show-up fee of \$5 in each of the three experiments, consistent with our standard procedures.[21] This is a constant across all subjects, and does not vary with the decisions the subjects faced. No subject faced losses.

The lower left panel of Figure 3 displays the elicited CRRA coefficients for our sample, based on a sample of 152 subjects. We employ the CRRA utility function introduced earlier to define the CRRA intervals represented by each row in the payoff matrix faced by the subject shown in Table 1, although other functional forms could also be used and would lead to similar conclusions.

---

[19] The reason for this design was to test for "order effects" in elicited risk attitudes, as reported in HJMR [2005b].

[20] The time between sessions for a given subject was usually one or two weeks. HJMR [2005a] show that elicited risk attitudes for this sample are stable over time horizons of several months.

[21] The subjects were recruited in lectures for experiments run during the usual lecture time and they received no show-up fee. In our case, subjects were recruited via Ex-Lab (http://exlab.bus.ucf.edu), consisting of a combination of e-mail alerts and online registration schedules from a subject pool database.

Each subject is assigned the midpoint of the CRRA interval at which they switch from choosing lottery A to lottery B.[22] The right column in Table 1 shows CRRA intervals associated with each switch point. The resulting distribution of risk attitudes is depicted in the bottom left panel in Figure 3. While a small portion of subjects appear to be risk-loving or risk-neutral, the bulk of the subjects appear to be averse to risk, with the modal response being in the neighborhood of a CRRA value of 0.5.

An alternative method for characterizing risk attitudes is an interval regression statistical model in which each subject's choice is the CRRA interval at which they "switch" from choosing lottery A to choosing lottery B. Using the predicted CRRA coefficients from the interval regression has a disadvantage: it throws out much of the individual variation that is not captured by socio-demographics. Thus, the fitted distribution of CRRA is smoothed, but the qualitative conclusions are unchanged. The advantages of using the fitted model are that, if reliable, the model allows us to predict risk attitudes for subjects without having to directly elicit them. It is costly and time consuming to have to run an elicitation task in addition to the test of the choice model of interest.[23]

In the interval regression,[24] we include a standard list of socio-demographic characteristics and dummy variables for each experimental session.[25] The estimates shown in the bottom right panel of Figure 3 are then obtained as predictions from this estimated model, setting each individual's characteristics equal to their actual values. Average CRRA is estimated to be 0.68 for this sample. The average standard error in the CRRA coefficient estimate was 0.14, and the 95%

[22] Several subjects switched two or more times. In this case we use the first and last switch points to define a relatively "fat" interval for that subject.

[23] In addition, there are problems asking a subject to give two "real responses" in the lab. First, there might be wealth effects, or expected wealth effects, when the earnings from one lottery affect valuations for the second lottery. Second, if one picks out one choice at random to pay the subject, one is assuming that one of the axioms of EUT (independence) is correct. If it is not, then this random payoff device can generate inconsistent preferences even if the underlying preferences are consistent. These points are well known in the experimental literature, and are important if one is attempting to identify which axioms of EUT might be in error.

[24] For subjects that participated in the 1x10x experiments, the data constitute a panel consisting of two observations for that subject, so we use panel interval regression models with random individual effects. We included a binary indicator in the regression to control for order effects when subjects did both 1x and 10x tasks.

[25] These were binary indicators for sex, race (black), a Business major, Sophomore status, Junior status, Senior status, high GPA, low GPA, Graduate student status, expectation of a post-graduate education, college education for the father of the subject, college education for the mother of the subject, and U.S. citizen status. We also included age in years.

confidence interval around the mean CRRA coefficient of 0.68 is between 0.41 and 0.96. Comparing the lower panels of Figure 3, the distribution of CRRA coefficients from the interval regression model (right panel) is more smoothed and concentrated around the mean relative to the distribution (left panel) that is obtained by directly eliciting the CRRA values.

How sensitive are our conclusions about the validity of EUT given the estimated width of the confidence interval above? Casual inspection of Figures 1 and 2 suggests there are wide differences in the CE over this range including the possibility that cost of making an EUT-inconsistent choice is nearly zero for some choice pairs. We examine the sensitivity of these conclusions more formally below.

3. Effects of CE Differences on Tests of EUT

We first consider the possibility that subjects may be more likely to choose inconsistently with EUT when the cost of doing so is trivial. Figure 4 shows the fraction of EUT consistent choices as a function of CE differences, using all the data from the two CR choices and the 6 PR choices. For each threshold listed on the bottom axis, the calculations underlying these figures drop any choice that entails a CE *difference* that is *less* than the indicated threshold. Thus, as the threshold gets above several pennies, many of the A*B* choices faced by risk averse subjects are naturally dropped from consideration. Figure 4 shows thresholds for the difference in CE varying from 0 cents up to 100 cents. The thin, dashed line shows the fraction of choices above the threshold on the bottom axis. Thus, for a threshold of 0 cents 100% of the choices are considered (i.e., the choices from the 6 PR choices plus the 2 CR choices, for all individuals). As the threshold increases, additional choices are dropped. Whether a choice is dropped depends on the estimated risk aversion of the subject and the parameters of the lotteries in each choice, since these are the factors determining the CE. The heavy, solid line shows the fraction of the remaining choices that are consistent with the EUT prediction.

Surprisingly, the ability of EUT to predict choices does not appear to increase as the threshold for CE differences is increased. Our earlier conclusion, that there is little support for EUT, is therefore not affected by excluding observations that were based on small differences in the CE of the lotteries (and where "small" is defined parametrically, so that the reader may individually decide what is small). Moreover, although not shown in Figure 4, we find that EUT does not do better than about 50% correctly predicted even if CE differences are required to exceed $3.00. Simple random chance would explain these data better than EUT.

Figure 5 undertakes the same analysis at the level of each individual task.[26] These results show that the fraction of choices above the threshold, shown by the thin dashed line, stays quite high for most of the preference reversal choice tasks. This is by design, given that we have a generally risk averse subject pool. By contrast, the fraction of choices above the threshold drops rapidly in the CR task involving lotteries A* and B*, as implied from Figure 1. In fact, given the CRRA values observed in our sample, no decisions have CE differences greater than about 30 cents for the A*B* pair. For this task and two of the preference reversal tasks, Pair 3 and 6, the fraction of choices consistent with EUT, shown by the heavy, solid line, does increase as the threshold increases. However, this only occurs for a small fraction of choices since the number of choices above the threshold falls rapidly for these three tasks. For the remaining tasks, there appears to be no relationship between the minimum threshold and the extent to which choices are consistent with EUT. This is particularly telling since these are the tasks for which a substantial fraction of choices exceed the threshold.

4. Allowing for Imprecision in Risk Elicitation.

We have seen that error rates do not decline even when CE differences are large. While we do not see any persuasive evidence that the size of CE differences affects our conclusions about EUT , we must recognize that our risk aversion coefficient estimates for individuals may be

[26] The lines in Figure 5 are defined identically to those in Figure 4.

imprecise. As seen in Figures 1 and 2, CE differences are very sensitive to the CRRA coefficient. Small changes in r can change an observed choice from being considered a trivial violation to a costly violation, or to no violation. Imprecision in estimating the CRRA coefficient must be taken into account when evaluating the data.

Imprecision may arise if our risk elicitation task does not yield precise estimates due to "trembling hand" error on behalf of the subject, or to our failure to elicit sufficient information to make more precise inferences about the risk attitudes of the subject. To illustrate an important but subtle point, imagine we had collected information on hair color and used that to explain the risk aversion choices of our subject. We anticipate that this would be a poor statistical model, generating extremely wide standard errors on our forecast of the individual's risk attitudes. As a result, it would be very easy to find a predicted risk aversion coefficient within a 95% confidence interval of the mean predicted value that includes the point of indifference. Thus one could almost always find a risk attitude that makes the observed choices consistent with EUT, but only because the statistical model was so poor. We have selected individual characteristics that are standard in empirical work of this kind, but there is always the risk that none of these characteristics help us predict risk attitudes carefully.

For all the analyses and tests employed so far, the way we characterize risk attitudes makes no difference to the conclusions we draw. Using the interval regression model to generate *average* risk aversion estimates for each individual yields indistinguishable results from the alternative, less parametric, approach which measures risk aversion for each individual using the observed interval at which the individual switches from safe to risky. The reason that these two approaches to inferring risk attitudes generate the same conclusions is that the *averages* from the interval closely approximate the *average* prediction from the interval regression model.

However, when considering the impact of the *precision* of the risk aversion estimates, the conclusions we draw are sensitive to how we statistically characterize risk attitudes. Thus, we will first consider the precision of raw responses, and then compare the precision of the CRRA coefficients estimated using interval regression models.

First, consider a minimally parametric approach that does not condition on the socio-demographic characteristics of the subjects. This allows us to focus on the imprecision inherent in the experimental task rather than prediction error in the regression model. Because subjects were only given ten questions in the risk aversion task, we only know the interval at which the subject switched from the safe to risky choice.[27] For each individual we know the upper and lower bounds of their "switching" interval. Any CRRA coefficient between these bounds is consistent with the observed switching behavior of the individual, and equally plausible *a priori*. Each CRRA coefficient in the interval is associated with a CE difference; hence, there is a range of equally plausible CE differences. For each individual and each choice, we pick the most "conservative" CRRA coefficient, that is, we pick the CRRA coefficient associated the smallest absolute value of the CE difference. Then, if the CE difference is below the chosen threshold, this observation is dropped. Thus, whenever it is plausible that the subject does not care about the choice given the bounds on the subject's risk aversion, that choice is excluded. The bottom panel of Figure 6 shows the results of this calculation. The horizontal axis again shows threshold values up to 100 cents and the thin dashed line shows the fraction of choices above the threshold. The darker line shows the fraction of EUT consistent choices when we allow for uncertainty over the precise CRRA coefficient for each individual. There do not appear to be conservative CRRA values for each subject, taking into account the interval nature of CRRA estimate, such that the predicted consistency of EUT rises much above 50%. For comparison, the top panel of Figure 6 shows the fraction of choices correctly predicted by EUT assuming *no* uncertainty in the risk aversion measure and using the midpoint of each individual's raw CRRA response interval. There is little difference in the fraction of EUT consistent choices between the top and bottom panels of Figure 6.

We also consider whether "trembling hand" errors in risk aversion could be driving these

[27] Of course, we could have asked more questions to "pin" the individual to a smaller interval. This alternative is implemented in a risk aversion elicitation task by Harrison, Lau, Rutström and Sullivan [2005].

apparent EUT violations. Suppose that the latent process that drives an individual's choices in the risk aversion experiment operates with some error, so that individuals may be observed switching earlier or later than their optimal switching points. To capture this idea, we expand the upper and lower bounds of the individual's observed CRRA interval to the midpoints of the adjacent intervals.[28] As above, we consider the range of CRRA values in this expanded interval, and then pick the that one that leads to the smallest CE difference in the same manner as before. The bottom panel of Figure 7 shows that EUT still preforms poorly even under this less exacting test (the top panel of Figure 6 is reproduced in the top panel of Figure 7 for comparison). Allowing for uncertainty over the risk aversion interval chosen does not provide any compelling new evidence in favor of EUT.

Now ask the same questions using the interval regression model to characterize risk attitudes. Figure 8 shows the effects of incorporating the forecast error of the model's prediction for each individual into the test of EUT. The top panel shows the fraction of choices that are both EUT consistent and above the CE threshold when the CRRA estimates are generated from the *average* of the prediction from the interval regression. In the bottom panel, forecast error from the regression is taken into account in a similar manner as described above for Figures 6 and 7. For each subject, we randomly draw a thousand CRRA estimates from the estimated distribution of CRRA values for that subject, in this case using the estimated mean and standard error of the forecast from the interval regression model as the estimated distribution.[29] We then use the CRRA estimate for the individual that generates the smallest of the absolute values of the CE difference between the two choices in each lottery pair.

---

[28] We do not extend the interval to the three intervals surrounding the chosen interval, since the trembling hand argument does not justify a uniform distribution over the "outer intervals." It simply says that somebody may have had a CRRA of 0.16 but chosen the interval with upper bound 0.15 since it was "close enough."

[29] The standard error of a forecast takes into account the uncertainty of the coefficients in the interval regression model. It is always larger than the standard error of the prediction, which assumes that those estimates are known exactly. These draws reflect the normal distribution appropriate for this estimated coefficient value, so 95% of the draws for each subject will be within ±1.96 standard errors of the point estimate. Thus, to emphasize, we are not allowing the CRRA values for any individual to take on values that are outside the realm of statistical precision given our experimental procedures.

Figure 8 shows the results of this calculation. What is striking here is that the fraction of choices that are above the threshold for CE differences drops to nothing when the threshold exceeds 10 cents. Hence, for each individual and each lottery, there exist "plausible" CRRA values such that the opportunity cost of an error under EUT is trivial.

Figures 6 and 8 pose a dilemma for the interpretation of the lottery choices from the perspective of EUT. One must decide which statistical characterization of risk attitudes is the best, in terms of reflecting the precision of inferences possible from our experimental procedure. Although there is nothing wrong with the interval regression characterization, we are firmly inclined towards the minimally parametric characterization since we have that for each individual in our sample. It makes fewer assumptions about the process generating the observed risk aversion choices, can be easily relaxed to undertake robustness checks as shown in Figure 7, and can be refined with simple extensions of the experimental procedures we used.[30] Thus, we conclude that *if one cannot directly elicit risk attitudes from the sample then EUT may be operationally meaningless since the estimated risk attitude coefficients suffer from too much imprecision*.

There are situations in which one might prefer the interval regression model, despite the relative imprecision of the estimates that result. Assume that the risk aversion test has been applied to a sample drawn from one population, and one wants to define a risk aversion distribution for use in interpreting data drawn from choices in a risk-sensitive task by a distinct sample drawn from the same population or a distinct population. All that one might know about the new sample are individual characteristics, such as sex and age. One could then generate conditional predictions for the new sample using the coefficient estimates from the interval regression model estimated on the first sample and the information on characteristics of the new, target sample. The minimally parametric characterization is not so attractive here, since it cannot be so easily conditioned on

[30] Harrison, Lau, Rutström and Sullivan [2005] consider two ways. One, noted earlier, is to "iterate" the MPL procedure several times so that subjects get 10 intervals *within* the interval they chose on the prior iteration. The other way in which one can tighten the CRRA interval is to administer the procedure several times over distinct lottery prizes, so as to span a more refined set of CRRA intervals.

individual characteristics. In many experimental situations considerations of cost may necessitate using predicted rather than elicited risk attitude coefficients. However, more work on specifying good predictive models is needed before such an approach can be meaningfully applied. Furthermore, there are many situations in which one only needs to know broad qualitative properties of the risk attitudes of subjects (e.g., are they risk-neutral or not), rather than precise estimates of degrees of risk aversion. For such purposes the within-sample procedures may be overkill.

5. Conclusions

We address the imprecision of the empirical parameterization of risk attitude in expected utility choice models and its impact on the probability of rejecting the underlying choice model. We provide an extended case study of the inferential problems that arise, assuming a CRRA form for the utility function. However, the issue is broadly generalizable to any situation in which parameters need to be estimated prior to testing the hypothesized choice function.[31] We adopt a procedure in which the risk attitude estimates are perturbed over successively wider intervals to provide a sense of the robustness of our conclusions regarding the hypothesized EUT choice function. We constrain the perturbations to intervals that are within estimated confidence intervals of the point estimates.

We begin by considering a context in which EUT appears to be a poor predictor of choice behavior. Under the null hypothesis of EUT and CRRA, we calculate the cost of choosing inconsistently with EUT conditional on estimated individual risk parameters. We find no evidence that the predictive power of EUT improves when we restrict the sample to choices that impose

[31] This need to account for the effect of errors that may arise in the elicitation task has been explicitly considered in the literature on using the "trade-off method" of Wakker and Deneffe [1996] to elicit the probability weighting function of rank-dependent utility theory. Using methods similar to the ones proposed here, Bleichrodt and Pinto [2000] use simulations to assess the robustness of their conclusions about the shape of the probability weighting function to subject errors. Even if errors cannot "propagate" in the elicitation task, any test of a choice theory that is formed conditional on a fitted parameter must take into account the precision with which that first stage parameter is estimated. More generally, the literature provides many examples in which predicted behavior is conditioned on risk attitudes, which then serve as a confound unless controlled for in some manner. For example, Cox, Smith and Walker [1985] and Harrison [1990] consider the effect of calibrating controls for risk attitudes on predicted bidding behavior in first-price sealed-bid auctions.

nontrivial costs on subjects. We proceed to examine two methods for estimating the first stage parameters, in this case individual risk parameters. Risk measures may be directly elicited by giving each subject a test, or may be predicted based on a statistical model that utilizes the information on subject risk response and demographics. In either case we find pervasive violations of the theory even when the opportunity costs of errors are substantial for a risk averse, expected utility maximizer. Furthermore, allowing for imprecision in our estimates due to "trembling hand" error demonstrates that we can estimate coefficients of relative risk aversion with sufficient precision to test EUT. Unfortunately, this is only true for the directly elicited "within" sample method. While the point estimates from the statistical model would lead to the same conclusion as when we directly elicit risk aversion measures, the imprecision of those estimates is such that they include CRRA values for which the cost of almost any error is negligible. While our qualitative conclusions about expected utility theory are unaffected by imprecision in measuring risk aversion, this concern is generally applicable to a wide variety of experimental situations.

Figure 1: Risk Attitudes and Common Ratio Tests of EUT

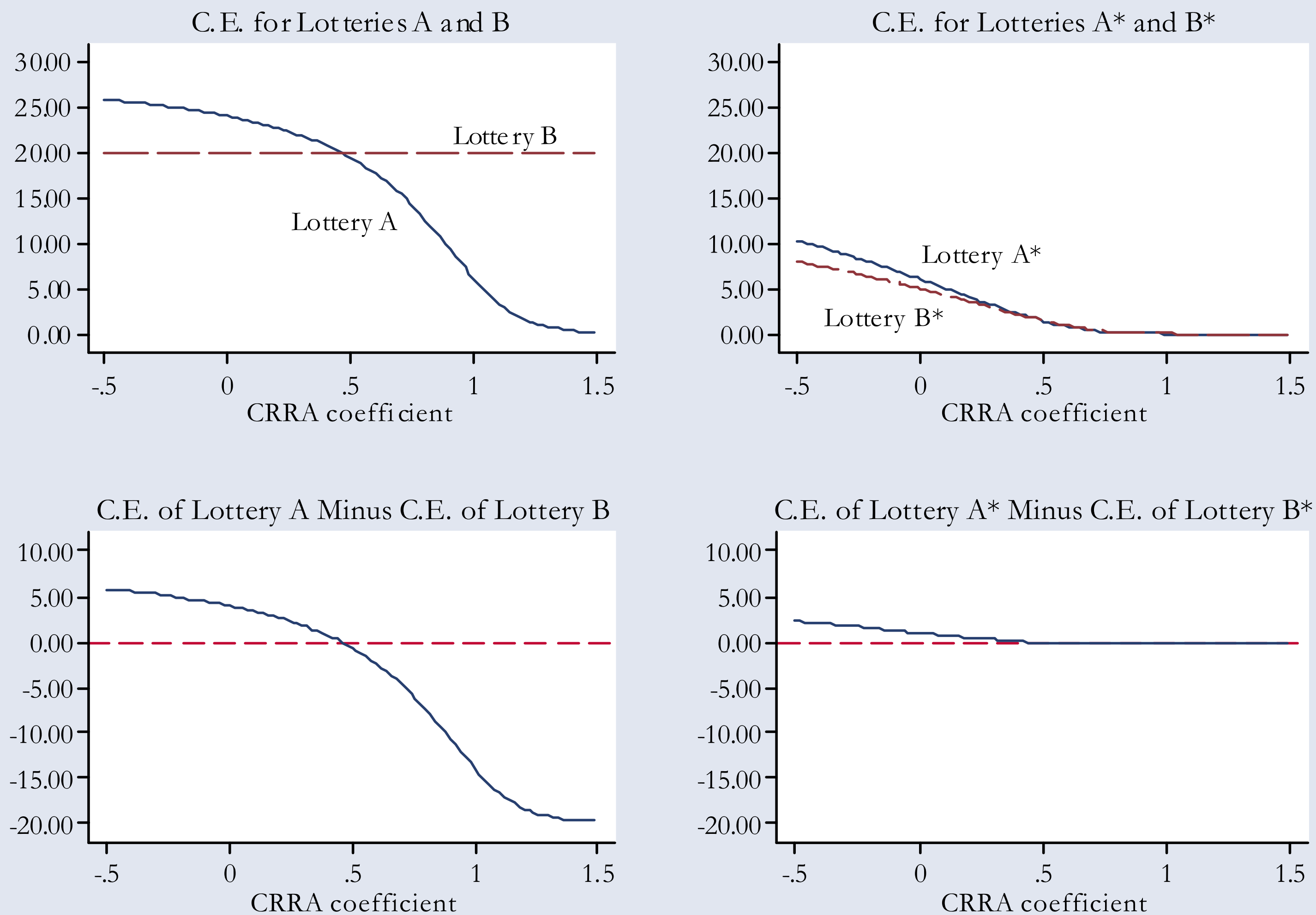


Figure 2: Risk Attitudes and Preference Reversal Choice Pairs

Difference in Certainty Equivalents Favoring P-Bet in Each Pair

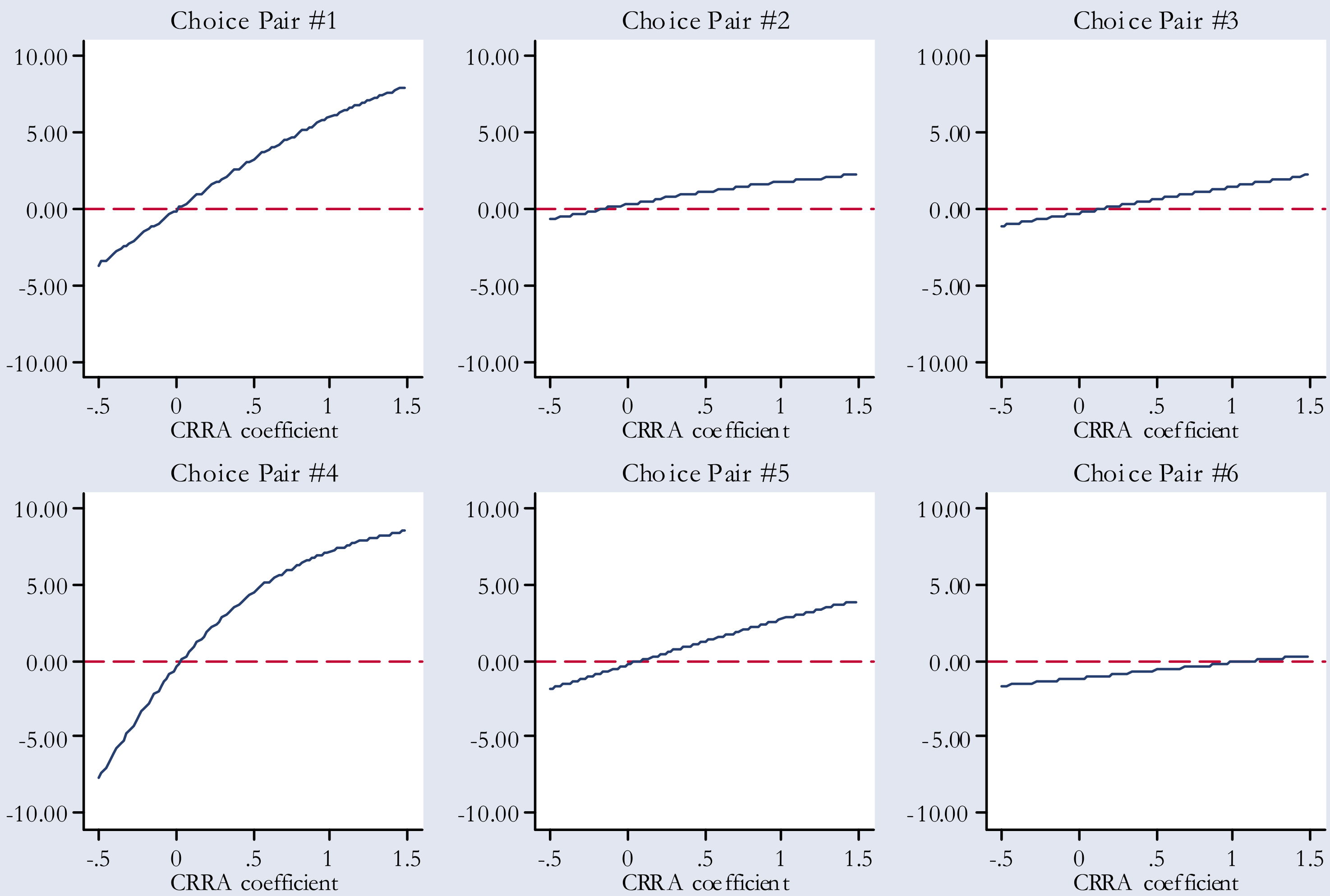


Table 1: Design of the Holt and Laury Risk Aversion Experiments

Standard Payoff Matrix

Lottery A		Lottery B		EV^A	EV^B	Difference in Expected Value	Range of CRRA coefficient *r* if last lottery A choice
Probability of Winning \$2	Probability of Winning \$1.60	Probability of Winning \$3.85	Probability of Winning \$0.10				
0.1	0.9	0.1	0.9	\$1.64	\$0.48	\$1.17	< -0.95
0.2	0.8	0.2	0.8	\$1.68	\$0.85	\$0.83	-0.95 < *r* < -0.49
0.3	0.7	0.3	0.7	\$1.72	\$1.23	\$0.49	-0.49 < *r* < -0.15
0.4	0.6	0.4	0.6	\$1.76	\$1.60	\$0.16	-0.15 < *r* < 0.15
0.5	0.5	0.5	0.5	\$1.80	\$1.98	-\$0.17	0.15 < *r* < 0.41
0.6	0.4	0.6	0.4	\$1.84	\$2.35	-\$0.51	0.41< *r* < 0.68
0.7	0.3	0.7	0.3	\$1.88	\$2.73	-\$0.84	0.68< *r* < 0.97
0.8	0.2	0.8	0.2	\$1.92	\$3.10	-\$1.18	0.97< *r* <1.37
0.9	0.1	0.9	0.1	\$1.96	\$3.48	-\$1.52	1.37< *r*
1	0	1	0	\$2.00	\$3.85	-\$1.85	

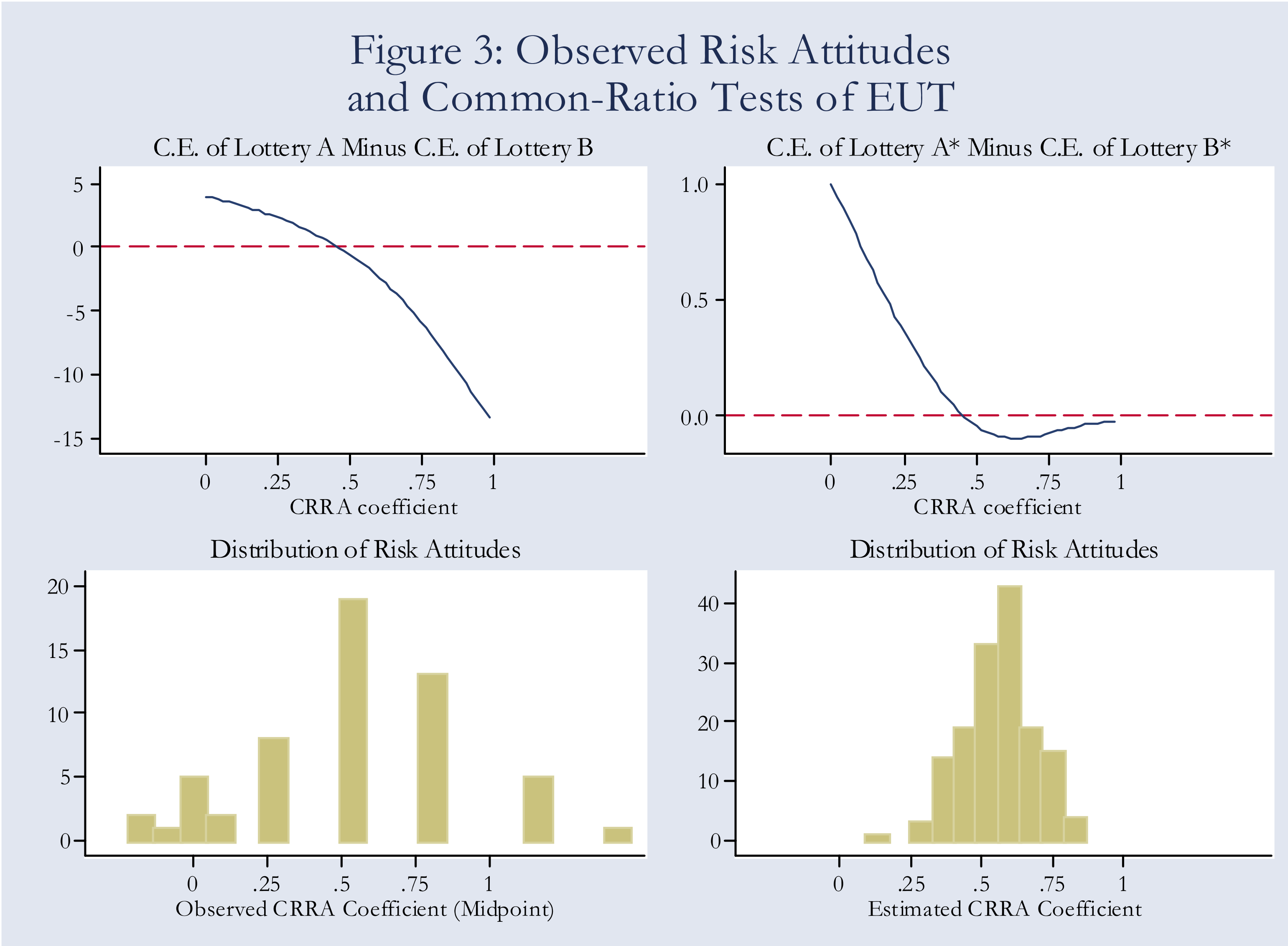


Figure 3: Observed Risk Attitudes
and Common-Ratio Tests of EUT
C.E. of Lottery A Minus C.E. of Lottery B
5
0
-5
-10
-15
0
.25
.5
.75
1
CRRA coefficient
C.E. of Lottery A* Minus C.E. of Lottery B*
1.0
0.5
0.0
0
.25
.5
.75
1
CRRA coefficient
Distribution of Risk Attitudes
20
15
10
5
0
0
.25
.5
.75
1
Observed CRRA Coefficient (Midpoint)
Distribution of Risk Attitudes
40
30
20
10
0
0
.25
.5
.75
1
Estimated CRRA Coefficient

Figure 4: Fraction of EUT Consistent Choices
as a Function of Certainty Equivalent Differences

All Choice Tasks

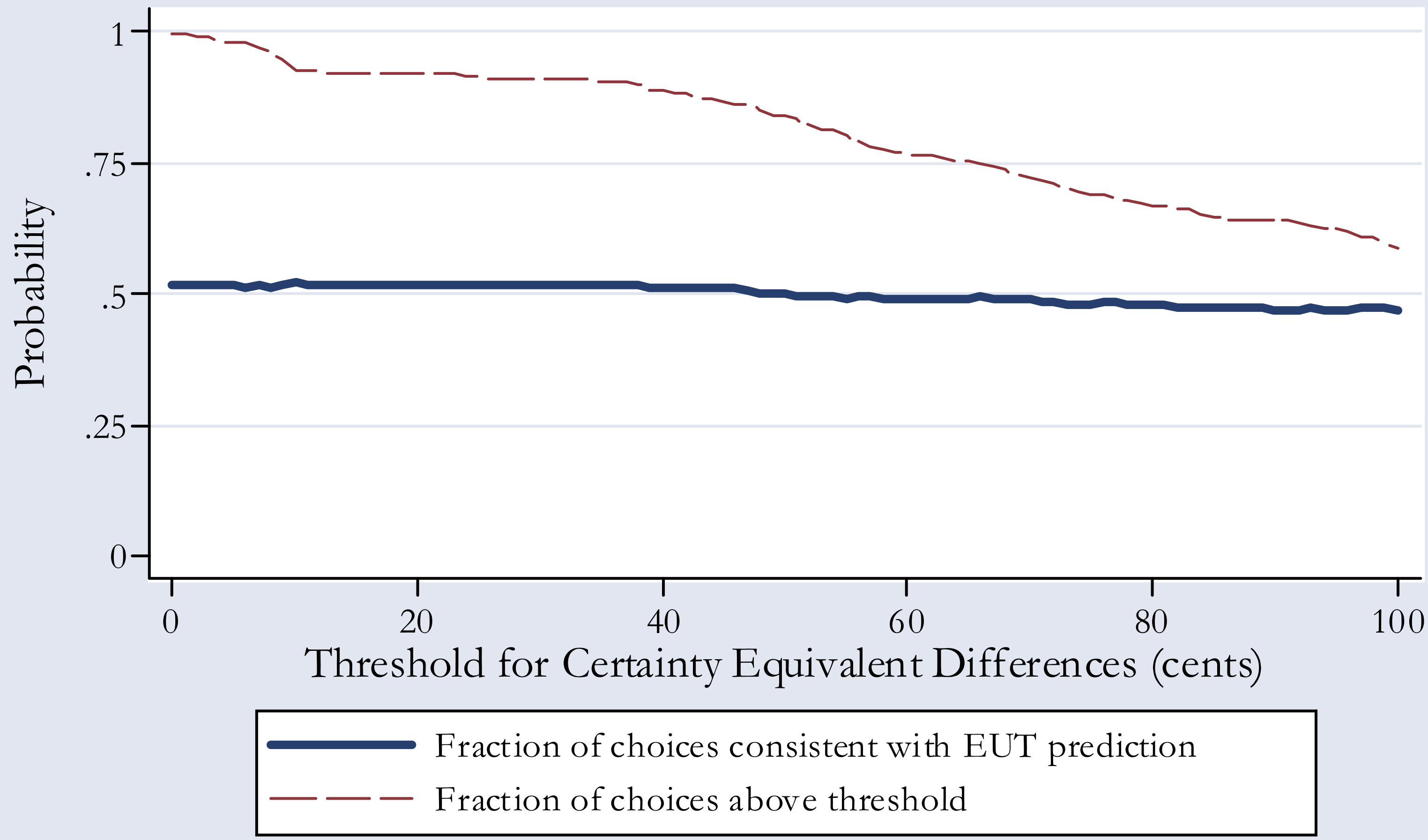
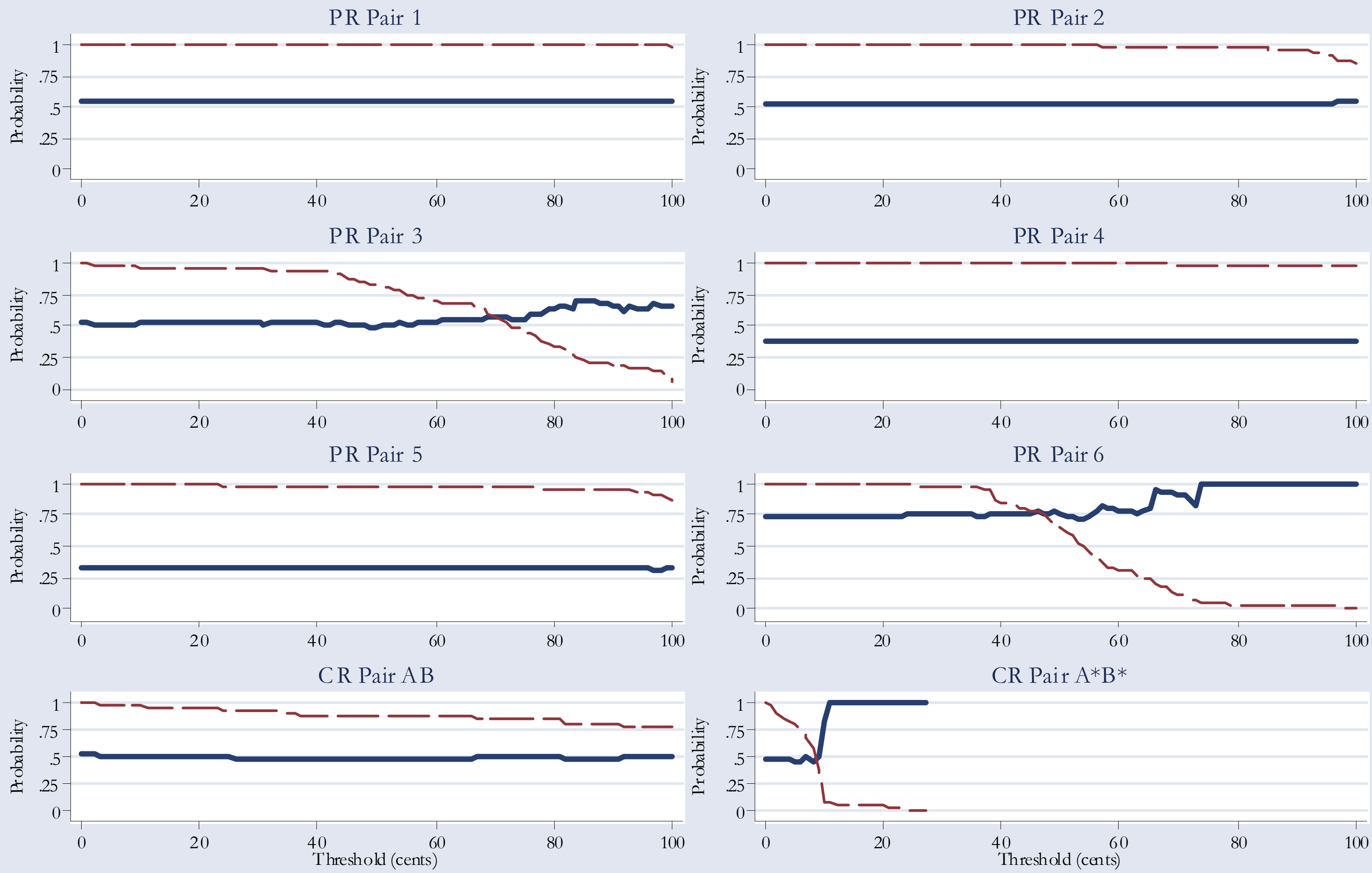


Figure 5: Fraction of EUT Consistent Choices as a Function of Certainty Equivalent Differences



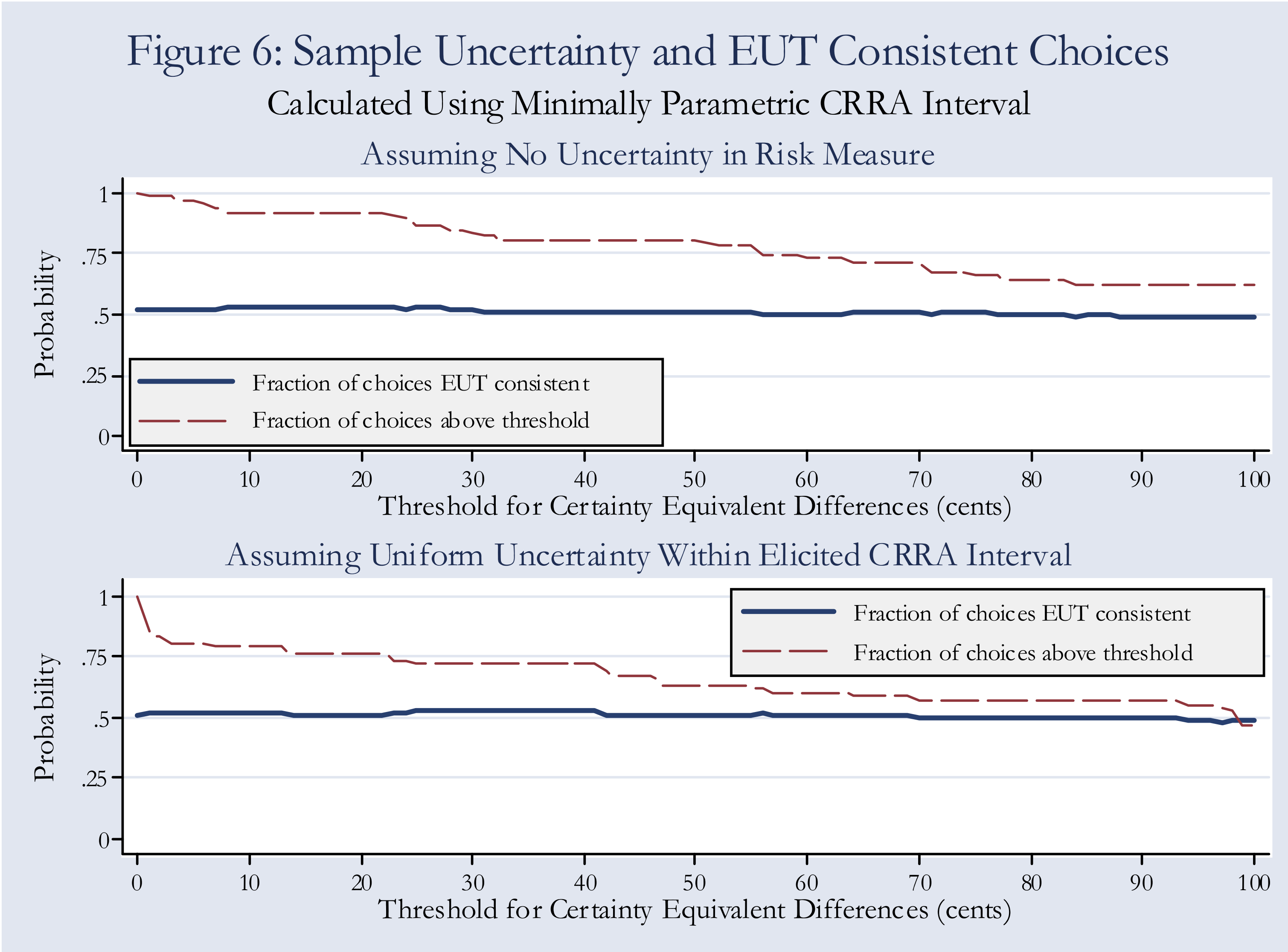


Figure 6: Sample Uncertainty and EUT Consistent Choices
Calculated Using Minimally Parametric CRRA Interval
Assuming No Uncertainty in Risk Measure
Probability
1
.75
.5
.25
0
Fraction of choices EUT consistent
Fraction of choices above threshold
0
10
20
30
40
50
60
70
80
90
100
Threshold for Certainty Equivalent Differences (cents)
Assuming Uniform Uncertainty Within Elicited CRRA Interval
Probability
1
.75
.5
.25
0
Fraction of choices EUT consistent
Fraction of choices above threshold
0
10
20
30
40
50
60
70
80
90
100
Threshold for Certainty Equivalent Differences (cents)

Figure 7: Additional Sample Uncertainty and EUT Consistent Choices

Calculated Using Minimally Parametric CRRA Interval

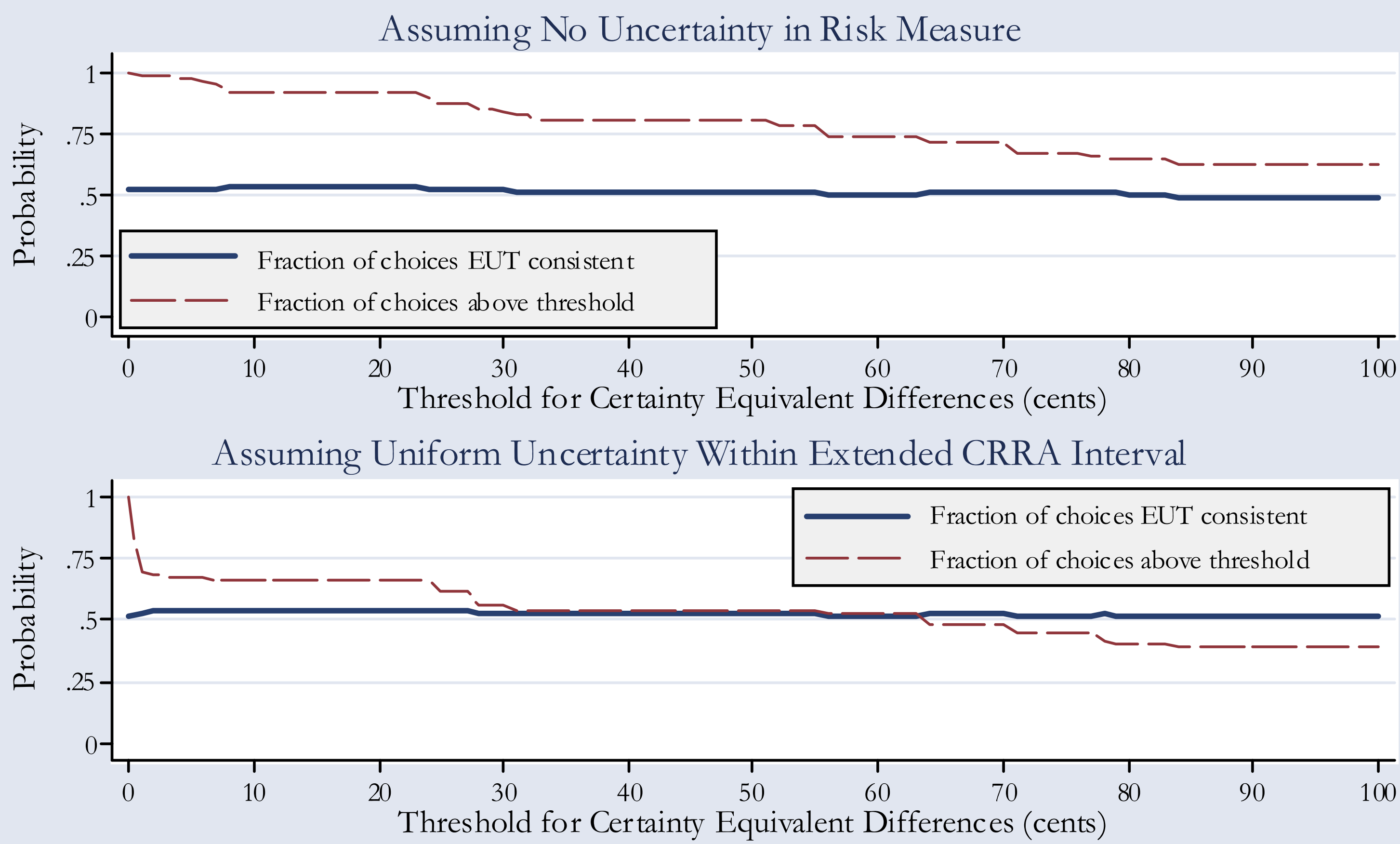
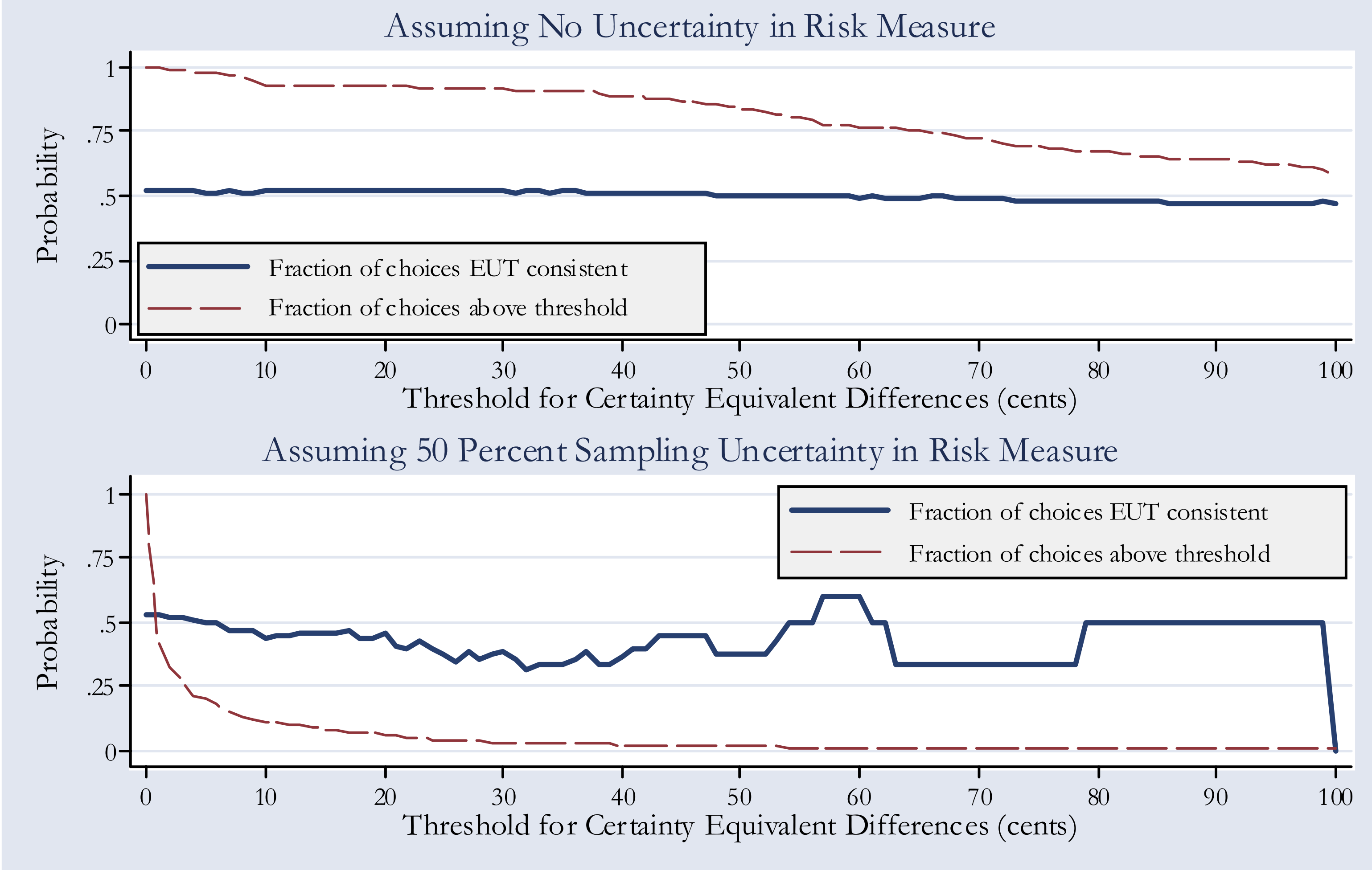


Figure 8: Sample Uncertainty and EUT Consistent Choices

Calculated Using CRRA Estimates from Interval Regression Model



References

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

Bleichrodt, Han, and Pinto, Jose L., "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis," *Management Science*, 46 (11), 2000, 1485-1496.

Botelho, Anabela; Harrison, Glenn W.; Pinto, Lígia M. Costa, and Rutström, Elisabet E., "Social Norms and Social Choice," *Working Paper 05-23*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Carbone, Enrica, "Investigation of Stochastic Preference Theory Using Experimental Data," *Economics Letters*, 57, 1997, 305-311.

Charness, Gary, and Rabin, Matthew, "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), August 2002, 817-869.

Coller, Maribeth; Harrison, Glenn W., and Rutström, Elisabet E., "Does Everyone Have Quasi-Hyperbolic Preferences?" *Working Paper 06-01*, Department of Economics, College of Business Administration, University of Central Florida, 2006.

Cox, James C., and Sadiraj, Vjollca, "Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory," *Games & Economic Behavior*, 56(1), July 2006, 45-60.

Cox, James C., "How To Identify Trust and Reciprocity," *Games and Economic Behavior*, 46(2), February 2004, 260-281.

Cox, James C.; Smith, Vernon L.; and Walker, James M., "Experimental Development of Sealed-Bid Auction Theory: Calibrating Controls for Risk Aversion," *American Economic Review (Papers and Proceedings)*, 75, May 1985, 160-165.

Engle-Warnick, Jim, "Inferring Decision Rules from Experimental Choice Data," *Working Paper*, Department of Economics, McGill University, May 2004.

Fudenberg, Drew, and Levine, David K., "A Dual Self Model of Impulse Control," *American Economic Review*, 96, 2006, forthcoming.

Grether, David M., and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69, 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.

Harless, David W., and Camerer, Colin F., "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62(6), November 1994, 1251-1289.

Harrison, Glenn W., "Risk Attitudes in First-Price Auction Experiments: A Bayesian Analysis," *Review of Economics and Statistics*, 82, August 1990, 541-546.

Harrison, Glenn W., "Field Experiments and Control," 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.; 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, 2003.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Temporal Stability of Estimates of Risk Aversion," *Applied Financial Economics Letters*, 1, 2005a, 31-35.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005b, 897-901.

Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," *Working Paper 05-01*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

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, 2005).

Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1013-1059.

Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility Theory *and* Prospect Theory: One Wedding and A Decent Funeral," *Working Paper 05-18*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Hey, John D., "Experimental Investigations of Errors in Decision Making Under Risk," *European Economic Review*, 39, 1995, 633-640.

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.

Karlan, Dean, and Zinman, Jonathan, "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Working Paper*, Department of Economics, Yale University, March 2005.

Kocher, Martin; Strauß, Sabine, and Sutter, Matthias, "Individual or Team Decision-Making: Causes and Consequences of Self-Selection," *Games & Economic Behavior*, 56(2), August 2006, 259-270.

Lazear, Edward P.; Malmendier, Ulrike, and Weber, Roberto A., "Sorting in Experiments, with Application to Social Experiments," *Working Paper*, Department of Economics, Stanford University, January 2006.

Leamer, Edward E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (New York: Wiley, 1978).

Loomes, Graham; Moffatt, Peter G., and Sugden, Robert, "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 2002, 103-130.

Loomes, Graham, and Sugden, Robert, "Incorporating a Stochastic Element Into Decision Theories," *European Economic Review*, 39, 1995, 641-648.

Loomes, Graham, and Sugden, Robert, "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 1998, 581-598.

Mayer, Thomas, "The Empirical Significance of Models," in M. Boumans (ed.), *Measurement in Economics: A Handbook* (San Diego, CA: Elsevier, 2007 forthcoming).

Magnus, Jan, R. "Local Sensitivity in Econometrics," in M. Boumans (ed.), *Measurement in Economics: A Handbook* (San Diego, CA: Elsevier, 2007 forthcoming).

Moffatt, Peter G., "Models of Decision and Choice," in M. Boumans (ed.), *Measurement in Economics: A Handbook* (San Diego, CA: Elsevier, 2007 forthcoming).

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

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

Rubinstein, Ariel, "Comments on the Risk and Time Preferences in Economics," *Unpublished Manuscript*, Department of Economics, Princeton University, 2002.

Wakker, P. P., and Deneffe, E., "Eliciting von Neumann-Morganstern Utilities When Probabilities are Distorted or Unknown," *Management Science*, 42, 1996, 1131-1150.