Temporal Stability of Estimates of Risk Aversion

by

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

September 2003

Abstract. Estimates of risk aversion can be obtained from controlled laboratory experiments. The temporal stability of those preferences is assumed in many applications. We test this assumption by eliciting risk aversion measures from subjects at two distinct times. We find evidence consistent with the stability assumption.

Key words: risk aversion, experimental economics

JEL classification codes: C81, C91, D81

†Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström), Department of Economics, Moore School of Business, University of South Carolina, USA (Johnson and McInnes). E-mail contacts: GHARRISON@BUS.UCF.EDU, EJOHNSON@MOORE.SC.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, NSF/MRI 9871019 and NSF/POWRE 9973669 to Rutström and grant NSF/SES 0213974 to McInnes. Corresponding author: Glenn Harrison, Department of Economics, College of Business Administration, University of Central Florida, Orlando, FL 32816-1400, USA. Supporting data and instructions are stored at http://www.bus.ucf.edu/gharrison/data/ee/ra2/.

Are laboratory estimates of risk aversion stable over time? If they are, then it may be possible to elicit measures from a sample of a given population and use them for later samples drawn from the same population, or at least interpret responses from the same individuals over time. If they are not stable, how significant is the instability? If risk attitudes are volatile, both with respect to the instrument used to elicit them or the passage of time, there is a serious problem for any researcher or policy maker who depends on historic estimates of risk attitudes for their analyses. We show that such concerns may be unfounded and demonstrate that, at least for the subject pool, elicitation instrument, and time span employed here, risk attitudes are stable.

We examine this issue using repeat observations from a sample of subjects that participated in a risk aversion elicitation experiment at two distinct points of time. Within the constraints of such a re-test, we conclude that the evidence is consistent with stability.

1. Estimating Risk Aversion in the Laboratory

Holt and Laury [2002] (HL) devise an elegant experimental measure for risk aversion using a multiple price list (MPL) 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 offers a 10% chance of receiving \$2 and a 90% chance of receiving \$1.60. The expected value of this lottery, EV^A, is \$1.64. Similarly, lottery B in the first row has a 10% chance of paying \$3.85 and 90% chance of paying \$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. Although the expected values and differences are shown in Table 1, they were not presented to subjects. As one proceeds down the matrix, the expected value of both lotteries increases, but the expected value of lottery B eventually becomes greater than the expected value of lottery A.

The subject makes a choice A or B in each row, and then one row is 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. 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. Option A is referred to as the "safe" choice while option B is the "risky" choice.

Each subject provided 10 responses in each of the two tests, since there are 10 rows in the table, and these data may be analyzed using a variety of statistical models. The responses can be reduced to a scalar if one looks at the *first* row in Table 1 at which the subject "switched" over from safe to risky.[1] This reduces the response to a scalar for each subject and task, but a scalar that takes on integer values between 0 and 10. Alternatively, one can use the well-known and widely used constant relative risk aversion (CRRA) utility function.[2] For each row of Table 1, one can calculate the implied bounds on the CRRA coefficient, and these are in fact reported by HL [2002; Table 3]. These intervals are shown in the final column of Table 1. Thus, for example, a subject that made 5 safe choices and then switched to the risky alternatives would have revealed a CRRA interval between 0.14 and 0.41, and a subject that made 7 safe choices would have revealed a CRRA interval between 0.68 and 0.97, and so on.

Harrison, Johnson, McInnes and Rutström [2003a][2003b] (HJMR) applied this basic procedure in a series of experiments conducted in late 2002 at the University of South Carolina. Their design was intended to examine the significance of "order" or "experience" effects in the original design of HL. In one treatment, called 1x10x, subjects participated in a risk aversion task with the payoffs shown in Table 1. They were then asked if they would like to give up their earnings in that task in order to play one more round in which payoffs were scaled up by a factor of 10. All subjects chose to do so. In the other treatment, a different sample of subjects drawn from the same population were simply given the latter risk aversion task where the payoffs were scaled by 10. This treatment serves as the obvious control for "order" or "experience" effects. We refer to this as the

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

[2] The CRRA utility function can be written as $u(x) = x^{(1-r)}/(1-r)$.

10x treatment. HJMR show both scale and order effects that are significant: the average predicted CRRA coefficient increased as the scale of the payoffs increased, but about half of that increase is attributed to an order effect.

To test whether these elicited risk aversion estimates were stable over time, we conducted a second round of experiments described below. We focus throughout on the responses of subjects to the task with payoffs scaled by a factor of 10.

2. The Re-Test Treatment

The initial sample of subjects participated in the risk aversion task between October 22, 2002 and November 20, 2002. There were 55 subjects in the 10x treatment and 123 subjects in the 1x10x treatment, or 178 in all. We recruited 31 of these subjects to a second experiment, in which we administered the 10x task only. There were two re-test sessions held on April 10 and May 9. Our analysis examines the stability of the responses of these 31 subjects over time, recognizing that we have a panel data set.

To gain an intuitive understanding of the economic significance of the retest effects, we estimate an interval regression model under the assumption of CRRA.[3] The dependant variable is the CRRA interval that subjects implicitly choose when they switch from lottery A to lottery B. This model is estimated on the responses based on the payoff scale of 10, so we have a panel data set of two observations for each subject. The model controls for unobserved individual effects by means of a random effects specification. A binary indicator to control for session effects in the re-test is also included, since there were two such sessions. The variables of interest for our purposes are binary indicator variables that indicate whether the response is from the first test or the re-test and whether the subject had previous experience with the 1x before answering the first 10x test, interacted with each other. Thus, we have four indicator variables, one for each quadrant in Table 2

[3] A small fraction switched two or more times. In this case we naturally use the first and last switch points to define a relatively "fat" interval for that subject.

below. We also estimate a second specification in which we add a standard list of socio-demographic characteristics.[4]

Table 2 shows the estimated coefficients on the variables of interest for the full model that includes demographic controls. Standard errors are given in parentheses. Omitting demographic controls does not change any of the results so we do not report these results separately. Comparing the coefficients going down the columns indicates the change in the estimated CRRA coefficient between the first test and the re-test given five to six months later. The second column shows the results for the subjects that had previously participated in the 1x10 treatment, and the third column shows the results for the subjects that had previously participated in the 10x treatment. We show these results separately because previous participation in a 1x task affects the quantitative value of elicited 10x CRRA in the first test, as explained in HJMR [2003a]. Subjects were about evenly divided between the previous 1x10x and 10x treatments. The p-values for differences in the coefficients are shown below the columns.

The effects for the subjects with no prior 1x experience is in many respects the "pure re-test" case because the exact same 10x task was applied in the test and re-test. The results are consistent with the elicited CRRA being stable: the difference in the coefficients is small in magnitude and statistically insignificant.

The effects for the subjects with the prior 1x10x experience appear, at first blush, to be consistent with there being some temporal instability of measured CRRA. The difference in the estimated treatment effects in Table 2 indicates that there is a significant decline in the CRRA coefficient during the re-test. However, this may be due to the effects of the prior experience with the 1x task on the estimated CRRA coefficient in the first test. We know from HJMR [2003a] that there is a significant order effect: subjects who did the 1x task had higher measured risk aversion for the subsequent 10x task than those who completed only the 10x task. Thus the response in the first

[4] 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.

test for these subjects is “artificially inflated” due to the order effect, and we hypothesize that the re-test response simply returns these subjects to a 10x CRRA level that is unaffected by an order effect.

To test these hypotheses, we can compare coefficient estimates across the rows. We expect to find an order effect in the second row of the table where we compare the responses in the first test across subjects who did and did not have prior experience with the 1x. As expected, we can reject the null hypothesis of no difference with a high degree of confidence. This confirms our previous results in HJMR [2003a] of a significant order effect. Our hypothesis, that this order effect does not persist, can be tested by looking across the third row, which compares the second round responses for these same groups. We find no significant differences in the coefficients here. Thus, the statistically significant change between the test and the re-test is due to the order effect, and not the re-test. In fact, if there had been *no* significant effect of the re-test we would have concluded that either (i) the order effect carries over into later sessions, or (ii) there was some temporal instability of the measure. Instead we find that the initial inflation of risk aversion due to order effects has worn off by the date of the re-test.

The distribution of predicted CRRA coefficients for this sample is shown in Figure 1, using the specification that includes the controls for demographics. The average predicted CRRA was 0.60 with a standard deviation of 0.27. These results are similar to our findings in HJMR [2003a] on a larger sample from the same population.

What is the statistical power of these measures of the effect of the re-test? Although the sample is 31, which might seem small in some settings, these are within subject tests. Moreover, the qualitative conclusions are generally robust to the addition of controls for session and demographic effects. Thus, we do not need to use up many degrees of freedom to draw the main qualitative conclusions. Nonetheless, we undertake a formal evaluation of the power of these data and this specification to detect certain re-test effects.

A power calculation measures our ability to correctly discern whether risk aversion has changed over the time that elapsed between the first test and the re-test. Because we find such a small difference in the coefficients between test and re-test in the no prior experience sample, it is obvious that enormous samples would be needed for our testing procedures to detect such a small difference. However, such a small difference (-0.009), even if true, would not have much economic significance. Given that our risk aversion measure is only designed to pin the estimated CRRA coefficient to fairly wide intervals, we can use the interval width of the narrowest interval as a benchmark. The narrowest interval, from Table 1, is for a CRRA coefficient between 0.41 and 0.68, which is only 0.27 wide. Given our minimum interval width, we can check for the power to detect a coefficient difference of 0.27. We find the power of the test exceeds 80% for the no prior experience sample and about 66% for the prior experience sample.[5]

3. Conclusions

Our results are consistent with risk aversion measures being stable over time. When subjects were given the same test separated by five to six months, we find no significant change in risk aversion. We also confirm earlier results from a much larger sample, that the order of tasks in a given session can affect risk aversion measures, but our findings here suggest those results will not persist over time.

It is therefore useful to state the constraints of our test, since we deliberately constructed a "minimalist test" of the stability hypothesis. We do not show if risk aversion measures elicited with other laboratory procedures are stable, and conjecture that certain procedures do not provide strong incentives or transparent tasks to subjects, perhaps encouraging instability.[6] We do not consider re-

[5] Data sets of 31 individuals, observed over two tests of their CRRA values, were simulated using the error term variances obtained from an initial calibration estimation. Each simulation was then repeated for 1000 iterations, allowing the recovery of the percentage of iterations for which the null hypothesis of no change in coefficients was correctly rejected (the power of the test). Additionally, in subsequent simulations, the coefficient values used to generate the data were varied from the base value of 0.27 to examine the sensitivity of the test power to alternative data generating processes.

[6] Isaac and James [2000] compare inferred risk aversion measures from in-sample comparisons of two different elicitation institutions, and find significant differences. One of the institutions they use is the Becker-DeGroot-Marschak (BDM) simulated auction, which elicits a willingness to pay that can be used to infer risk attitudes, following Harrison [1986]. The other institution they use is a first-price (FP) sealed bid auction, from which one can infer risk attitudes if one is willing to assume non-cooperative Bayesian Nash

tests for longer periods of time, such as years. In part this is deliberate, in the spirit of giving the stability hypothesis a minimal test. 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. Thus it would be plausible *a priori* to see risk aversion change with major life events, such as the birth of children or marriage, but for it to remain stable conditional on controls for those states. The payoffs in our task range from $1 up to $38.50. Although this is a useful range from the perspective of experimental applications of a measure of risk aversion, our results say nothing by themselves about risk aversion over wider domains or in the loss domain.

References

Harrison, Glenn W., "An Experimental Test for Risk Aversion," *Economics Letters*, 21(1), 1986, 7-11.

Harrison, Glenn W., "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 82, December 1992, 1426-1443.

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

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Individual Choice in the Laboratory: Paradox Reloaded," *Unpublished Manuscript*, Department of Economics, College of Business Administration, University of Central Florida, 2003a; available at http://www.bus.ucf.edu/gharrison/wp/.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Isaac, R. Mark, and James, Duncan, "Just Who Are You Calling Risk Averse?" *Journal of Risk and Uncertainty*, 20(2), 2000, 177-187.

Equilibrium bidding behavior. The BDM institution is known to have weak incentive properties (Harrison [1992]), and the FP institution requires that one maintain strong assumptions about equilibrium behavior in order to infer risk attitudes. Our test is simpler and more direct, employing in-sample responses with the same elicitation institution. Moreover, the logic of truthful revelation of preferences is much easier for subjects to understand in our institution than in the BDM procedure, and does not depend on Nash Equilibrium assumptions.

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	-∞, -0.95
0.2	$2	0.8	$1.60	0.2	$3.85	0.8	$0.10	$1.68	$0.85	$0.83	-∞, -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, ∞

<u>Note</u>: 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: Estimated Treatment Effects from Interval Regression Estimation Results

Model with demographic controls

Standard errors in parentheses

	Prior experience with 1x	No prior experience with 1x	P-value for differences in coefficients
First test	0.968 (0.423)	0.603 (0.381)	0.015
Re-test	0.750 (0.422)	0.594 (0.384)	0.282
P-values for differences in coefficients	0.013	0.915	

Note: Demographic controls were included in the statistical analysis underlying these treatment effects. 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.

Figure 1: Distribution of Risk Attitudes for Re-Test Subjects

Panel Interval Regression Model Estimated on
10x Responses of 31 Subjects (N=62)