Johannes Maier; Maximilian Rüger:

Measuring Risk Aversion Model-Independently

Measuring Risk Aversion Model-Independently

Johannes Maier[*] and Maximilian Rüger[†]

October 11, 2010

Abstract

We propose a new method to elicit individuals' risk preferences. Similar to Holt and Laury (2002), we use a simple multiple price-list format. However, our method is based on a general notion of increasing risk, which allows classifying individuals as more or less risk-averse without assuming a specific utility framework. In a laboratory experiment we compare both methods. Each classifies individuals almost identically as risk-averse, -neutral, or -seeking. However, classifications of individuals as more or less risk-averse differ substantially. Moreover, our approach yields higher measures of risk aversion, and only with our method these measures are robust toward increasing stakes.

Keywords: Risk Aversion, Multiple Price-List, Elicitation, Laboratory Experiment, Holt and Laury Method, Mean Preserving Spreads, Non-EUT, Increasing Risk.

JEL Classification Numbers: D81, C91.

1 Introduction

In order to measure individuals' risk attitudes, the multiple price-list method of Holt and Laury (2002) has become the industry standard in experimental economics. Major advantages that led to the popularity of the Holt and Laury (HL) tables include its transparency to subjects (easy to explain and implement), its incentivized elicitation, and that it can be easily attached to other experiments where risk aversion may have an influence. Nevertheless, the HL method has also its drawbacks. The major disadvantage is that it requires a specific utility framework such as expected utility theory (EUT) in order to classify subjects as more or less risk-averse.[1] If individuals' risk preferences are heterogeneous in the way that some act according to EUT while others rather act according to non-EUT, it becomes problematic to use the HL tables in order to classify subjects' risk attitudes. The reason is that the HL method is not based on a general notion of increasing risk which is satisfied by EUT *and* non-EUT models.

To account for this disadvantage, we propose a modification of the HL tables. This new method is based on the well-known increasing risk definitions of Rothschild and Stiglitz (1970). These definitions are solely in terms of attributes of the distribution function and are therefore independent of a utility framework. Moreover, they have been used to characterize risk aversion in both EUT and non-EUT models (e.g. Machina, 1982; or Chew, Karni and Safra, 1987). By just imposing 'duality' asserting that less risk-averse individuals accept riskier gambles (see Diamond and Stiglitz, 1974), our method enables us to classify subjects as more or less risk-averse without assuming a specific utility framework. It is therefore applicable with heterogeneous risk preferences. Furthermore, our approach is more robust toward probability weighting, since it uses variations in outcomes (i.e. mean preserving spreads) and holds probabilities of outcomes constant at 50%.

In a laboratory experiment we directly compare the HL method and our method using low and high stakes. We find that both methods yield the same classification of individuals concerning the *direction* of risk attitudes (i.e. risk-averse, risk-neutral, or risk-seeking). However, we also find that both methods yield diverging results concerning the *intensity* of risk attitude. The classification of individuals as being more or less risk-averse than others is quite different between both methods. Moreover, we find that our method yields higher levels of risk aversion intensity that are much closer to what is observed in the field. These estimates of risk aversion intensity are robust toward multiplying the stakes by five only when our method is used. For the HL method we can confirm the result of Holt and Laury (2002) that increasing the stakes increases risk aversion. It is also shown that these results

---

[1]Holt and Laury (2002) use specific parametric forms of EUT in order to classify subjects.

are robust toward possible confounds like certain switching preferences or order effects.

The paper is structured as follows. In section 2 we review the HL tables and note further advantages and disadvantages. We then propose a new method that shares the advantages but not the disadvantages of the HL tables in section 3. The experiment used to directly compare both methods is explained in section 4. The results of our experiment as well as their robustness are discussed in section 5 and the conclusion can be found in section 6. In appendix A we generalize our new method so that it can be used for different purposes of measuring the intensity of risk attitude.

2   The Holt and Laury Method

Measuring the intensity of risk preferences is very important for theoretical predictions. Also, in experiments individuals' decisions are often (partly) driven by their risk preferences. In order to control for that, the multiple price-list method of Holt and Laury (2002) is commonly used in experiments nowadays. Table 1 presents the original HL design.

Table 1: The Holt and Laury Method

Row No.	Option A		Option B		RRA if row was last choice of A and below all B	$EV[\text{A}]-$ $EV[\text{B}]$	$Var[\text{A}]-$ $Var[\text{B}]$
	Outcome A1 = \$2.00	Outcome A2 = \$1.60	Outcome B1 = \$3.85	Outcome B2 = \$0.10			
1	Prob. 1/10	Prob. 9/10	Prob. 1/10	Prob. 9/10	$[-1.71 \ ; \ -0.95]$	1.17	-1.25
2	Prob. 2/10	Prob. 8/10	Prob. 2/10	Prob. 8/10	$[-0.95 \ ; \ -0.49]$	0.83	-2.22
3	Prob. 3/10	Prob. 7/10	Prob. 3/10	Prob. 7/10	$[-0.49 \ ; \ -0.14]$	0.50	-2.92
4	Prob. 4/10	Prob. 6/10	Prob. 4/10	Prob. 6/10	$[-0.14 \ ; \ 0.15]$	0.16	-3.34
5	Prob. 5/10	Prob. 5/10	Prob. 5/10	Prob. 5/10	$[0.15 \ ; \ 0.41]$	-0.18	-3.84
6	Prob. 6/10	Prob. 4/10	Prob. 6/10	Prob. 4/10	$[0.41 \ ; \ 0.68]$	-0.51	-3.34
7	Prob. 7/10	Prob. 3/10	Prob. 7/10	Prob. 3/10	$[0.68 \ ; \ 0.97]$	-0.85	-2.92
8	Prob. 8/10	Prob. 2/10	Prob. 8/10	Prob. 2/10	$[0.97 \ ; \ 1.37]$	-1.18	-2.22
9	Prob. 9/10	Prob. 1/10	Prob. 9/10	Prob. 1/10	$[1.37 \ ; \ \infty)$	-1.52	-1.25
10	Prob. 10/10	Prob. 0/10	Prob. 10/10	Prob. 0/10	non-monotone	-1.85	0.00

An individual makes a decision between option A and option B in each of the ten rows. Option A as well as option B can have two different outcomes (A1 or A2 and B1 or B2) with varying probabilities over the ten rows. The expected outcome of option A is higher for the first four rows and lower for the last six rows (as indicated by the second to last column in Table 1). So, a risk-neutral subject should choose option A in row 1 to 4 and then switch over and choose option B in row 5 to 10. However, as option B has a higher variance (indicated by the last column in Table 1), there is a trade-off when to switch to option B. Clearly, by row 10 everybody should have switched to option B as it yields the higher outcome with certainty. An individual who switches to option B between row 6 and row 10 is classified as being risk-averse and the more risk-averse individual will switch later as she needs a higher expected value to choose the more variable option. Someone who switches earlier to option B (between row 1 and row 4) is classified as risk-seeking by similar

arguments. In column 6 of Table 1 we report the risk preference intensity measured by the amount of relative risk aversion (RRA) that is induced from the switching behavior if we assume the class of constant relative risk-averse (CRRA) utility functions.[2]

The advantages of the HL method are due to its design. It is very easy to explain to subjects since they only have to choose between option A and option B in each row.[3] It is incentivized and usually one of the ten rows is randomly selected and paid out for real. And because it is so easy to implement, the HL table can be attached to other experiments where risk aversion may play a role.

Nevertheless, the HL method also has its disadvantages. One disadvantage is that there is no flexibility in adjusting the ranges of RRA without affecting the round-numbered probabilities. So, for instance, if one would want to decrease the ranges of the RRA intervals in row 4 to 6 in order to better classify most subjects' risk attitudes (according to Holt and Laury (2002) 75% of the subjects fall into this category), one would have to give up the round-numbered probabilities in Table 1. One way to circumvent this problem is proposed by Andersen et al. (2006) using a complex more-stage procedure and thereby loosing the advantages of the simple HL design mentioned above.

The use of variations in probabilities (whereas outcomes are held constant) makes the HL tables sensitive to probability weighting. For instance, by using the standard parametric Prospect Theory assumptions (Tversky and Kahneman, 1992) on the probability weighting function, we obtain the result that a subject with a linear utility function should choose A only for the first three and not for the first four rows. Such an individual would be classified as risk-seeking in HL. Of course, this makes it difficult to draw conclusions about the shape of the utility function.

The major disadvantage of the HL tables is, however, that they are not based on a general notion of increasing risk. They need a specific utility framework, namely EUT, in order to classify subjects as more or less risk-averse.[4] However, evidence rather suggests that risk preferences are heterogeneous and subjects follow different models of risky choice.[5] It is then

---

[2] Note that the bound in rows 3 and 4 of $r = -0.15$ as reported in the original article of Holt and Laury (2002) is in fact according to our calculation $r = -0.14$. Also, if the subject chooses always option B, his relative risk aversion is $r \in (-\infty; -1.71]$.

[3] A variant of the design is to induce a single switching point, i.e. to ask for the row where the subjects would want to switch from A to B.

[4] In order to discriminate between intensities of risk aversion, HL use EUT and the specific class of CRRA functions. Using the HL method and assuming for instance Tversky and Kahneman's (1992) Prospect Theory would require a 'trade-off' between the curvatures of the utility function and the probability weighting function since both simultaneously influence the level of risk aversion. However, even if such a 'trade-off' could be made, there is no way to compare subjects in case they follow different models of risky choice.

[5] The evidence that many individuals behave according to non-EUT models is vast. For instance, a recent study by Harrison, Humphrey and Verschoor (2010) finds a 50/50 share of EUT and Prospect Theory.

not only problematic to impose EUT in order to classify subjects, but also to assume the very same choice model over all subjects. Hence, a more *general* measure of risk aversion intensity is needed that allows for a classification across different underlying models. In the next section, we therefore propose a modification of the HL method that is based on a general 'behavioral' notion of risk aversion, namely an aversion to mean preserving spreads.

3 A Model-Independent Method

In this section we propose a new method that shares the advantages of the HL table but not its disadvantages as mentioned above. Table 2 presents our new approach.

Table 2: Our Elicitation Method

Row No.	Option A		Option B		RRA if row was first choice of A and above all B	RRA if row was last choice of A and below all B
	Prob. 1/2 Outcome A1	Prob. 1/2 Outcome A2	Prob. 1/2 Outcome B1	Prob. 1/2 Outcome B2		
1	0.05	4.95	2.65	2.75		$[-0.51 \; ; \; -0.13]$
2	1.10	3.90	2.65	2.75		$(-\infty \; ; \; -0.51]$
3	**2.40**	**2.60**	**2.65**	**2.75**		non-monotone
4	2.40	2.60	2.00	3.40	$[2.27 \; ; \; \infty)$	
5	2.40	2.60	1.90	3.50	$[1.70 \; ; \; 2.27]$	
6	2.40	2.60	1.75	3.65	$[1.18 \; ; \; 1.70]$	
7	2.40	2.60	1.60	3.80	$[0.86 \; ; \; 1.18]$	
8	2.40	2.60	1.45	3.95	$[0.65 \; ; \; 0.86]$	
9	2.40	2.60	1.05	4.35	$[0.36 \; ; \; 0.65]$	
10	2.40	2.60	0.20	5.20	$[0.13 \; ; \; 0.36]$	

Again, subjects choose in each row between option A and option B. As in the HL table, option A as well as option B has two possible outcomes. However, instead of varying the probabilities and keeping the outcomes constant over all rows as in the HL table, we rather vary the outcomes and keep the probabilities constant (i.e. all probabilities are equal, namely 50%). First note that an individual with monotone preferences will always prefer option B over option A in row 3 of Table 2 (this is similar to row 10 in Table 1) as here option B first-order stochastically dominates (or more specifically, state-wise dominates) option A.

We now compare options in row 4 to those in row 3. While option A is identical to the one in row 3, option B in row 4 is a mean preserving spread of the one in row 3. We can therefore say that option B becomes more risky in the sense of the very general increasing risk definition of Rothschild and Stiglitz (1970), while option A stays the same. In row 5 option A is again unaltered whereas option B is a further mean preserving spread of the one in row 4 and thus a further increase in risk. This continues until row 10. By just imposing 'duality' stating that less risk-averse individuals should take riskier gambles, we can say that someone (call her j) who preferred option B in the first four rows and option A in the last six rows is more risk-averse than someone (call her i) who preferred option B in the first five rows and option A in the last five rows. Such a statement can be made without referring to

any particular utility framework. The property of 'duality' is based on the concept of mean utility preserving spreads first proposed by Diamond and Stiglitz (1974).[6] To see how, note that there exists a hypothetical gamble that is a mean preserving spread of option B in row 4, a mean preserving contraction of B in row 5, and leaves j just indifferent between A and B. This hypothetical gamble then is a mean utility preserving spread of A for j. At this point i still prefers B, so her hypothetical gamble representing a mean preserving spread of B in row 5, a mean preserving contraction of B in row 6, and leaving i just indifferent between A and B is a mean preserving spread of j's hypothetical gamble. It follows that j is more risk-averse than i. To illustrate how our table relates to the one of Holt and Laury (2002), we state in the last two columns of Table 2 how our method would elicit measures of RRA if we would also assume CRRA.

Risk seeking is identified through switches of choices in the first two rows of Table 2. Consider again the options in row 3, but now compare them to those in row 2. Now the 'less attractive' option A is altered by a mean preserving spread when going from row 3 to row 2, while option B stays the same. Only a very risk-seeking individual would like this spread so much that she would now prefer option A in row 2. In row 1 option A is a further mean preserving spread. Now, also less extreme risk seekers, who in row 2 were still choosing option B, are lured by the further increase in risk toward choosing option A in row 1. An individual who is risk-neutral, or is very close to being risk-neutral, will always choose option B in Table 2 since its expected value is higher than the one of option A in all rows.

Both options in Table 2, option A and option B, are always risky. This avoids the 'certainty effect', a well-known problem of any elicitation method using certainty equivalents.[7] The concept of riskiness we use in our table follows the established theoretical literature. "Clearly riskiness is related to dispersion, so a good riskiness measure should be monotonic with respect to second-order stochastic dominance. Less well understood, perhaps, is that riskiness should also relate to location and thus be monotonic with respect to first-order stochastic dominance, in particular, that a gamble that is sure to yield more than another should be considered less risky. Both stochastic dominance criteria are uncontroversial [...]" (Aumann and Serrano, 2008, p811)

In Table 2 we use both criteria. Option B first-order stochastically dominates option A in row 3 and can therefore be considered less risky. Going downward from row 3 option A

---

[6] Diamond and Stiglitz (1974) study this concept solely within EUT. By referring to simple compensated spreads, Machina (1982) and Chew, Karni, and Safra (1987) used it to analyze risk aversion in non-EUT models.

[7] "The certainty effect introduces systematic errors into any method based on certainty equivalents." (McCord and de Neufville, 1986, p57) It describes the widely observed phenomenon that certain alternatives are perceived in a fundamentally different way then risky alternatives, even if the risk is negligible.

stays unaltered whereas option B gets worse in terms of second-order stochastic dominance. Going upward from row 3 option A gets worse in terms of second-order stochastic dominance whereas option B stays the same. Individuals who switch from option B to A after row 3 are risk-averse (the earlier the more risk-averse). And individuals who switch from option A to option B before row 3 are risk-seeking (the later the more risk-seeking).[8]

Although the two concepts we use for our elicitation method, mean preserving spreads and mean utility preserving spreads, were originally analyzed within EUT, it has become common in other models as well to understand risk aversion in terms of this more behavioral definition, namely as an aversion to mean preserving spreads. Subsequently, it is this definition that is used when risk aversion is analyzed in non-EUT models (see e.g. Machina, 1982; Chew, Karni and Safra, 1987; or Schmidt and Zank, 2008). And as Machina (2008, p80) notes, most non-EUT models "are capable of exhibiting first-order stochastic dominance preference, [and] risk aversion [. . .]." This shows that our method can classify subjects as more or less risk-averse across various models of risky choice.

Using variations of outcomes (i.e. mean preserving spreads) not only makes it easy for subjects to compute expected values but also allows us quite some flexibility in designing the range of the intervals to elicit estimates of relative risk aversion if we adopt the CRRA framework of Holt and Laury (2002). In principle, this could also be achieved in the HL table, but only at the price of stating odd probabilities. By contrast, in our table probabilities stay always at 50% and only outcomes vary. We believe that subjects are more experienced in dealing with odd outcomes (such as price tags) than with odd probabilities.[9] More importantly, constant probabilities of 1/2 are much less sensitive to probability weighting. Already Quiggin (1982) uses 1/2 as plausible fixed point in his theory. "The claim that the probabilities of 50-50 bets will not be subjectively distorted seems reasonable, and [. . .] has proved a satisfactory basis for practical work [. . .]." (Quiggin, 1982, p328) This becomes important when decisions taken in the table are interpreted solely in terms of the curvature of a utility function.

---

[8]Note that mean preserving contractions of option A (option B) could be applied in addition to the mean preserving spreads of option B (option A) when going downward (upward) from row 3. This variant of the design could be useful if one wanted to induce similar changes in both options over all rows of our table. However, since it may come at the cost of increased complexity for the subjects, we chose not to do so.

[9]In appendix A we provide a more general treatment of our method. This shows how our method can be easily modified to meet different requirements on the elicitation ranges.

4 The Experiment

The experiment was computer-based and was conducted at the experimental laboratory MELESSA of the University of Munich. It used the experimental software z-Tree (Fischbacher, 2007) and the organizational software Orsee (Greiner, 2004). 232 subjects (graduate students were excluded) participated in 10 sessions and earned 11 euros (including 4 euros show-up fee) on average (with a maximum (minimum) of 30 (4.10) euros) for a duration of approximately one hour.

In the beginning of the experiment subjects received written instructions that were read privately by them. At the end of these instructions they had to answer test questions that showed whether everything was understood. There was no time limit for the instructions and subjects had the opportunity to ask questions in private. The experiment started on the computer screen only after everybody answered the test questions correctly and there were no further questions.

The further procedure of the experiment was the following. Each subject made decisions in four tables.[10] Again, they could take as much time as they wanted in order to make their decisions. After all subjects had made their decisions, an experimental instructor came to each subject to let them randomly determine their payoff from the tables.[11] Before they saw what their payoff from the experiment was they could again see how they actually decided in the randomly determined relevant table. At the end of the experiment all subjects further answered a questionnaire about their socio-economic characteristics. As soon as everybody had answered the questionnaire they were payed in private (not by the experimenter) and could leave.

Table 3: Our Adjusted Elicitation Method

Row No.	Option A		Option B		RRA if row was first choice of A and above all B	RRA if row was last choice of A and below all B
	Prob. 1/2 Outcome A1	Prob. 1/2 Outcome A2	Prob. 1/2 Outcome B1	Prob. 1/2 Outcome B2		
1	0.03	4.89	2.62	2.72		$[-0.49 \; ; \; -0.14]$
2	1.01	3.91	2.62	2.72		$[-0.96 \; ; \; -0.49]$
3	1.40	3.52	2.62	2.72		$[-1.70 \; ; \; -0.96]$
4	1.65	3.27	2.62	2.72		$(-\infty \; ; \; -1.70]$
5	**2.36**	**2.56**	**2.62**	**2.72**		non-monotone
6	2.36	2.56	1.77	3.57	$[1.37 \; ; \; \infty)$	
7	2.36	2.56	1.61	3.73	$[0.97 \; ; \; 1.37]$	
8	2.36	2.56	1.42	3.92	$[0.68 \; ; \; 0.97]$	
9	2.36	2.56	1.09	4.25	$[0.41 \; ; \; 0.68]$	
10	2.36	2.56	0.26	5.08	$[0.15 \; ; \; 0.41]$	

---

[10]Eight treatments varied which tables in which order a subject received. The treatments are further explained below.

[11]Each subject had to role four dice. First, a four-sided die determined which of the four tables was payoff-relevant. Second, a ten-sided die determined which row in the payoff-relevant table was selected. And lastly, two ten-sided dice determined whether the amount A1 or A2 (if A was chosen in the relevant table and row) or whether the amount B1 or B2 (if B was chosen in the relevant table and row) was payed out to them (in addition to the show-up fee of 4 euros).

As mentioned above, each subject made decisions in four tables. Two of the tables were low-stakes tables and two of them were high-stakes tables, where all low-stake outcomes were multiplied by five. In total, we used eight different tables. One of them was the original HL table (HLol) as outlined in section 2 (Table 1) and another was the original HL table but with all outcomes multiplied by five (HLoh). In order to being able to directly compare the HL method and our method, we adjusted our tables to the exact same ranges of RRA that were used by Holt and Laury (2002). A third table therefore used our method but adjusted to the original low-stake outcomes of the HL table (MRal) as outlined in Table 3. And a fourth table used our method adjusted to the high-stake version of the original HL table (MRah), where all outcomes in Table 3 are multiplied by five.

Subjects further received our table (Table 2) from section 3 (MRol). In designing this table we employed criteria mentioned by Holt and Laury (2002). There is an approximately symmetric range of RRA around 0, 0.5, 1, and 2. Based on the experimental results of Holt and Laury (2002), our table has only two risk seeking ranges and therefore more ranges for reasonable degrees of risk aversion. There was also a high-stakes version of this table where all outcomes are multiplied by five (MRoh). Again, in order to directly compare both methods, we also adjusted the HL tables to the exact same ranges of RRA that were used in our tables. Table 4 shows the adjusted HL table for low stakes (HLal). Again, the high-stakes version of Table 4 (HLah) multiplied all outcomes by five.

Table 4: The Adjusted Holt and Laury Method

Row No.	Option A		Option B		RRA if row was last choice of A and below all B
	Outcome A1 = $2.00	Outcome A2 = $1.60	Outcome B1 = $3.85	Outcome B2 = $0.10	
1	Prob. 29/100	Prob. 71/100	Prob. 29/100	Prob. 71/100	$[-0.53 \; ; \; -0.14]$
2	Prob. 40/100	Prob. 60/100	Prob. 40/100	Prob. 60/100	$[-0.14 \; ; \; 0.12]$
3	Prob. 49/100	Prob. 51/100	Prob. 49/100	Prob. 51/100	$[0.12 \; ; \; 0.36]$
4	Prob. 58/100	Prob. 42/100	Prob. 58/100	Prob. 42/100	$[0.36 \; ; \; 0.65]$
5	Prob. 69/100	Prob. 31/100	Prob. 69/100	Prob. 31/100	$[0.65 \; ; \; 0.85]$
6	Prob. 76/100	Prob. 24/100	Prob. 76/100	Prob. 24/100	$[0.85 \; ; \; 1.19]$
7	Prob. 86/100	Prob. 14/100	Prob. 86/100	Prob. 14/100	$[1.19 \; ; \; 1.70]$
8	Prob. 95/100	Prob. 5/100	Prob. 95/100	Prob. 5/100	$[1.70 \; ; \; 2.37]$
9	Prob. 99/100	Prob. 1/100	Prob. 99/100	Prob. 1/100	$[2.37 \; ; \; \infty)$
10	Prob. 100/100	Prob. 0/100	Prob. 100/100	Prob. 0/100	non-monotone

Each of all eight different tables was received by 116 subjects and all 232 subjects were in either of eight different treatments. The treatments were designed to control for order effects, not only whether subjects answered low- or high-stakes tables first, but also whether HL tables or our tables (adjusted and original) were answered first. The eight treatments ensured that every subject had the same ex-ante expected income.[12]

---

[12]The eight treatments were: 1. HLol, MRal, HLoh, MRah; 2. MRol, HLal, MRoh, HLah; 3. MRal, HLol, MRah, HLoh; 4. HLal, MRol, HLah, MRoh; 5. HLoh, MRah, HLol, MRal; 6. MRoh, HLah, MRol, HLal; 7. MRah, HLoh, MRal, HLol; 8. HLah, MRoh, HLal, MRol.

In the comparison of the HL method and our method we will ask several questions. Firstly, whether both methods yield the same classification of individuals concerning both the direction and the intensity of risk attitude. Secondly, whether both methods yield similar levels of risk aversion intensity. Thirdly, what the effect is of increasing the stakes (i.e. multiplying all outcomes by five) on these RRA estimates. And lastly, how robust our results are.

5 Results

5.1 Directions of Risk Attitudes

Before analyzing the intensities of risk attitudes, we can ask how many of the subjects that can be classified are risk-averse, risk-neutral, and risk-seeking. Under low stakes, we find that 79% are risk-averse, 11% are risk-neutral, and 10% are risk-seeking in the HL tables. The respective numbers for our tables are 81%, 9%, and 10%. Under high stakes, 88% are risk-averse, 7% are risk-neutral, and 5% are risk-seeking in the HL tables whereas the respective numbers are 88%, 6%, and 6% in our tables. This suggests that both methods yield identical classifications of subjects concerning the *direction* of risk attitude.

This finding is confirmed when we look at within subject classifications. Among the subjects that can be classified, we find that 84% have the same direction of risk attitude across both methods under low stakes, where 76% are risk-averse, 4% are risk-neutral, and 4% are risk-seeking in both methods. Under high stakes, 90% have the same direction of risk attitude in both methods with 84% being risk-averse, 4% risk-neutral, and 2% risk-seeking.

For each relevant comparison of both methods (i.e. HLol vs. MRal, HLoh vs. MRah, HLal vs. MRol, and HLah vs. MRoh) we perform sign tests in order to see whether the HL tables yield a different classification on the direction of risk attitude than our tables. We find that none of the comparisons is significant.[13] We can now state our first result.

Result 1 *The HL method and our method do not yield a different classification of individuals concerning the direction of risk attitude (risk-averse, risk-neutral, or risk-seeking).*

5.2 Intensities of Risk Attitudes

While Result 1 shows that both methods yield the same classification concerning the *direction* of risk attitude, another question is whether both methods also yield the same

---

[13]The results (p-values, number of observations N) of two-sided Sign tests are: HLol vs. MRal ($p = 1.0000$, $N = 76$), HLoh vs. MRah ($p = 1.0000$, $N = 71$), HLal vs. MRol ($p = 0.1094$, $N = 66$), and HLah vs. MRoh ($p = 1.0000$, $N = 71$).

classification concerning the *intensity* of risk attitude. Are individuals similarly classified as more or less risk-averse across both methods? This is an important question as it answers whether the methodological problem of using the HL method (due to the required assumption of EUT) is indeed an empirically relevant one. If both methods yielded the same classification of subjects concerning the intensity of risk attitude, results of experimental studies that used the HL method (and thereby assumed EUT) in order to classify subjects as more or less risk-averse would not be flawed. If, however, subjects are differently classified as more or less risk-averse in both methods, the methodological problem of assuming EUT for the classification in the HL method is in fact empirically relevant.

In order to answer this question we take the following approach. We first take a subject in a HL table and determine whether she is classified as more, less, or equally risk-averse than any other subject. We then take the same subject in our table and determine as well whether she is classified as more, less, or equally risk-averse than each of the other subjects. When we now compare this subjects pairwise comparisons (to all other subjects) in both tables, we can identify for this subject whether she is classified the same or differently (against each of the other subjects) across both tables. Thus, we know against how many of the other subjects this subject changes her risk aversion ranking across both tables. Since we perform this task not only for one specific but for all subjects, we can identify against how many of the other subjects a subject changes her ranking across methods on average. When we compare the HL tables to our tables we find that on average subjects change their ranking to 49% of the other subjects.[14] So, the average subject has a different ranking toward almost half of the other subjects across methods.[15]

Result 2 *The HL method and our method yield a different classification of individuals concerning the intensity of risk aversion relative to other individuals.*

Results 1 and 2 are of great interest for other experimental studies that use the HL method

---

[14]The comparisons we make yield the following results. The average subject changes her ranking (of being more, less, or equally risk-averse) to 54%, 52%, 47%, and 42% of the other subjects in the respective comparisons of HLol vs. MRal, HLoh vs. MRah, HLal vs. MRol, and HLah vs. MRoh.

[15]Note that subjects also change their ranking to other subjects across stakes (but within methods). However, we can test whether subjects change their ranking across methods more than across stakes. Holding the stakes effect constant, we find that subjects increase their ranking changes to other subjects on average by 36% when the method in addition to the stakes changes. When performing two-sided Wilcoxon signed-rank sum tests we find that subjects change their ranking significantly more across methods than across stakes (HLol vs. HLoh against HLol vs. MRah ($z = -5.262$, $p = 0.0000$, $N = 68$), HLal vs. HLah against HLal vs. MRoh ($z = -6.366$, $p = 0.0000$, $N = 67$), MRol vs. MRoh against MRol vs. HLah ($z = -2.431$, $p = 0.0151$, $N = 63$), and MRal vs. MRah against MRal vs. HLoh ($z = -3.169$, $p = 0.0015$, $N = 57$)). Results do not change when using two-sided Mann Whitney U tests instead (HLol vs. HLoh against HLol vs. MRah ($p = 0.0000$, $N = 170$), HLal vs. HLah against HLal vs. MRoh ($p = 0.0000$, $N = 157$), MRol vs. MRoh against MRol vs. HLah ($p = 0.0273$, $N = 135$), and MRal vs. MRah against MRal vs. HLoh ($p = 0.0000$, $N = 133$)).

Figure 1: Cumulative Distributions of RRA for All Tables
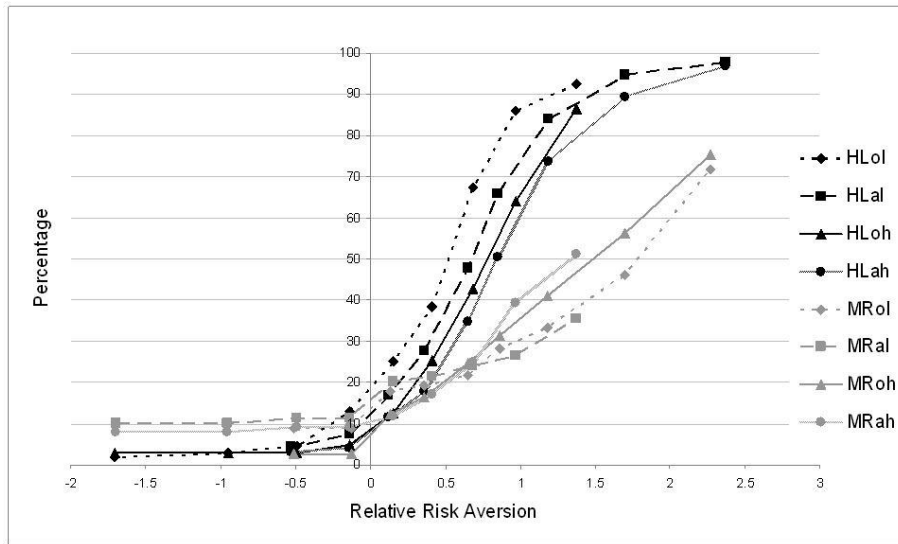
to control for risk aversion in observed behavior but which are not specifically interested in the absolute level of risk aversion intensity. However, there are other studies where the absolute level of risk aversion is important in order to derive quantitative predictions of a theoretical model. In the following we therefore investigate what the levels of risk aversion intensity are in both methods and how robust these RRA estimates are toward increasing the stakes.

Concerning the level of the intensity of risk attitude we again find systematic differences between both methods. Figure 1 shows the cumulative distributions of RRA for all eight different elicitation tables.[16] The cumulative distributions of relative risk aversion of all four HL tables lie above those of our four tables. While almost none of the subjects lies in the highest RRA range in the HL tables, many subjects fall into the highest RRA range when our method is used.[17] The medians of RRA using the HL method are all below the medians when our method is used. In the HL tables the medians lie in RRA ranges below one (HLol: $[0.41; 0.68]$; HLal: $[0.65; 0.85]$; HLoh: $[0.68; 0.97]$; HLah: $[0.65; 0.85]$) but they lie in RRA ranges above one with our tables (MRol: $[1.70; 2.27]$; MRal: $[1.37; \infty)$; MRoh: $[1.18; 1.70]$; and MRah: $[0.97; 1.37]$). While Figure 1 shows the overall picture, Figures 2 and 3 show the specific distributions that need to be compared in analyzing our data.[18]

Figure 2 compares the HL method and our method. For low and high stakes we can

---

[16]We used uniform distributions within the RRA ranges.

[17]Note that as the highest range goes to infinity, the cumulative distribution functions do not end at 100 for the displayed values of RRA. Similarly, as the lowest range goes to minus infinity, the cumulative distribution functions do not start at 0 for the reported RRA values. This is also the reason why we cannot investigate the means of RRA but only the medians.

[18]In Figures 2 and 3 we omitted naming the axis, but since all distribution lines are taken from Figure 1 and just considered in isolation, the notation of Figures 2 and 3 is of course the same as in Figure 1.

Figure 2: Cumulative Distributions of RRA: Comparing Methods
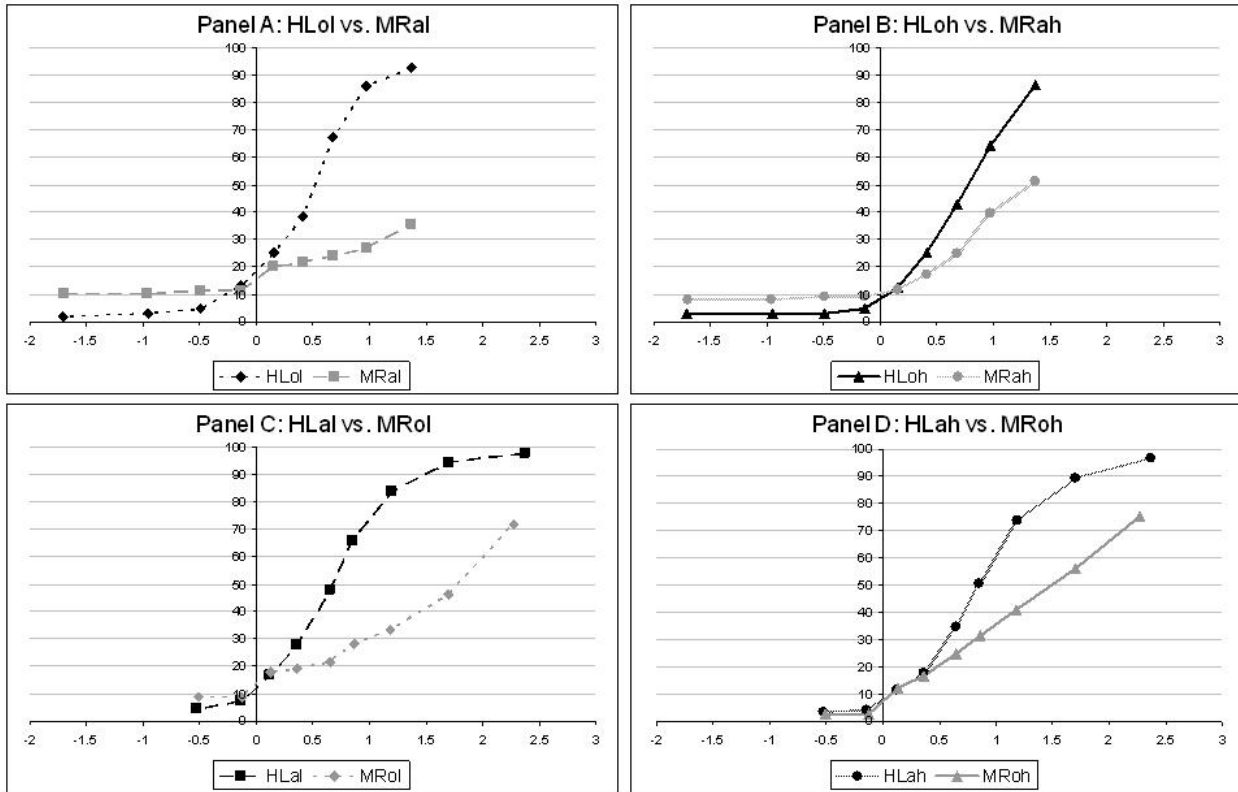
compare the original to the adjusted tables since the adjustment was such that the ranges of RRA were identical in both tables. In all four panels the HL method yields clearly lower measures of RRA. This is the case no matter whether the adjustment took place for the HL method (Panels 2C and 2D) or for our method (Panels 2A and 2B) or whether we look at low (Panels 2A and 2C) or high stakes (Panels 2B and 2D).

Figure 3 shows the effect of increasing the stakes in both methods. Since the cumulative distributions of the high-stakes HL tables lie below those of the low-stakes HL tables (Panels 3A and 3C), increasing the stakes seems to increase relative risk aversion. This is in contrast to our method where increasing the stakes does not cause risk aversion to increase. The cumulative distributions of our high-stakes tables rather cross those of our low-stakes tables (Panels 3B and 3D). This picture seems not to be affected by the fact whether we compare original (Panels 3A and 3B) or adjusted tables (Panels 3C and 3D).

Since each subject made decisions in four tables, we also test these differences using matched pairs. Table 5 relates the original HL tables to our adjusted tables, such that RRA ranges are identical and can be directly compared. And Table 6 relates the adjusted HL tables to our original tables such that all comparisons have identical RRA intervals. Reported are the number of observations (N), the z-values, and the p-values of two-sided

Figure 3: Cumulative Distributions of RRA: Comparing Stakes
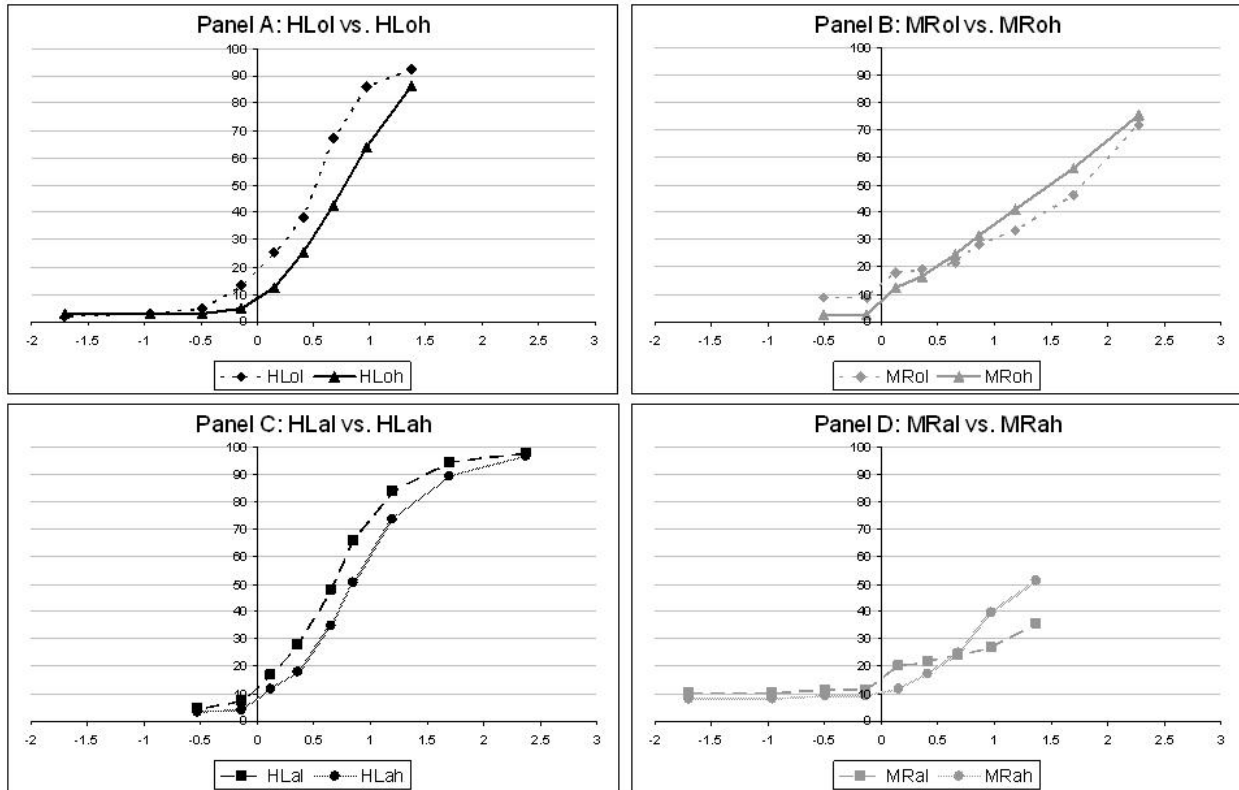
Wilcoxon signed-rank sum tests.[19]   In both tables, Table 5 and Table 6, in each cell it is tested if the elicitation method to the left yields different measures of RRA than the elicitation method above. Consider an example from Table 5. If HLol is compared to MRal, then a z-value of $z = -4.922$ indicates that the left table (HLol) yields lower measures of RRA than the right table (MRal).

Comparing first the HL method and our method, we observe significantly higher measures of RRA when our method is used (HLol vs. MRal and HLoh vs. MRah in Table 5; and HLal vs. MRol and HLah vs. MRoh in Table 6). This holds at the 1%-level for all four comparisons of Figure 2. We can therefore state the following result.

Result 3 *Our method yields a higher intensity of risk aversion than the HL method.*

Looking at the effect of increasing the stakes, we see that there is a significantly positive

---

[19]Our results do not change when using two-sided Kolmogorov-Smirnov tests instead. The p-values are all $p = 0.000$ for the comparisons of Figure 2 (with $N = 186, 179, 172, 168$ respectively for Panels 2A, 2B, 2C, 2D). Thus, compared to the HL method our method yields significantly higher measures of RRA. For the comparisons of Figure 3, we get $p = 0.798$ ($N = 151$) and $p = 0.227$ ($N = 155$) for Panels 3B and 3D, respectively. And we get $p = 0.002$ ($N = 210$) and $p = 0.167$ ($N = 189$) for the respective Panels 3A and 3B. Hence, while measures of RRA are not significantly different between low and high stakes when our method is used, this is not the case with the HL method. Here, increasing the stakes significantly increases measures of RRA in the original (two-sided at 1%-level) as well as in the adjusted (one-sided at 10%-level) HL tables.

Table 5: Wilcoxon signed-rank sum tests (HLo and MRa)

(N) z-value p-value		MRal	MRah
MRal			(61) 0.488 / 0.6257
HLoh		(72) -4.000 / 0.0001	(71) -2.580 / 0.0099
HLol	(99) -5.081 / 0.0000	(76) -4.922 / 0.0000	(71) -4.815 / 0.0000

effect on the RRA measure in the HL tables (HLol vs. HLoh in Table 5; and HLal vs. HLah in Table 6). In contrast, there is no such effect observed when our method is used (MRal vs. MRah in Table 5; and MRol vs. MRoh in Table 6). So, for the comparisons of Figure 3, we see that increasing the stakes by a factor of five increases risk aversion significantly at the 1%-level only when the HL method is used. With our method, there is no significant effect of increasing stakes.[20]

Result 4 *While increasing the stakes by a factor of five increases the intensity of risk aversion with the HL method, increasing the stakes has no effect on the intensity of risk aversion with our method.*

Table 6: Wilcoxon signed-rank sum tests (HLa and MRo)

(N) z-value p-value		MRol	MRoh
MRol			(65) -0.270 / 0.7875
HLah		(70) -3.215 / 0.0013	(71) -4.293 / 0.0000
HLal	(90) -4.957 / 0.0000	(66) -3.756 / 0.0002	(67) -4.654 / 0.0000

Results 3 and 4 show that our method not only yields higher risk aversion estimates than the HL method, but also that our estimates are robust toward multiplying all outcomes by five (thereby indicating CRRA). This is not the case for the HL estimates. Here, we find increasing relative risk aversion (IRRA). Our findings for the HL method are completely in line with the findings of Holt and Laury (2002). Nevertheless, the results for our method are much closer to what is observed in the empirical literature. Several empirical studies

---

[20]As can be seen from Tables 5 and 6, even the sign of the z-value is positive (adjusted tables) *and* negative (original tables).

indicate a measure of RRA roughly between 1 and 2 (e.g. Tobin and Dolde, 1971; Friend and Blume, 1975; Kydland and Prescott, 1982; Hildreth and Knowles, 1982; Szpiro, 1986; Chetty, 2006; or Bombardini and Trebbi, 2010) and Mehra (2003, p59) notes that "most studies indicate a value for α that is close to 2."[21] An experimental study by Levy (1994) rejects the existence of IRRA. Other empirical studies (e.g. by Szpiro, 1986; Friend and Blume, 1975; Brunnermeier and Nagel, 2008; or Calvet and Sodini, 2010) find supportive evidence for CRRA. Fehr-Duda et al. (2010) show in their experimental study that IRRA is entirely driven by transformations of the probability weighting function as stakes increase. In contrast to the HL method, our method is invariant to probability weighting. Hence, this may be the reason why there is a stakes effect only with the HL method.

Another possible explanation for the results of Holt and Laury (2002) and ours is that subjects need higher incentives when they have to exert more cognitive effort. In the HL tables, subjects need to exert more cognitive effort than in our tables since the varying probabilities are difficult to handle. By contrast, in our tables all probabilities are one half and 50/50 odds seem easy to work with. When subjects have too little incentives to exert the cognitive effort that is required to reveal their true level of risk aversion, it seems reasonable that they anchor their decision on the 50/50 choice (i.e. row 5 in the original HL tables).[22] Our results and the results of Holt and Laury (2002) suggest that increasing the stakes increases risk aversion with the HL method. So, as incentives increase subjects are more willing to exert such effort and thereby show their true level of risk aversion.[23] This might be the reason that Holt and Laury (2002) do not observe such a stake effect with hypothetical payoffs. With our method, subjects do not need much cognitive effort to handle 50/50 chances. Already our low-stakes tables give sufficient incentives to show subjects' true level of risk aversion. Increasing the stakes has therefore no effect on risk aversion with our method. It is interesting to note that increasing the stakes by a factor of five seemed not to be sufficient to show subjects' true level of risk aversion in the HL tables since risk aversion was still lower than in our tables. This is, however, consistent with the findings of Holt and Laury (2002) who increased stakes by factors of 20, 50 and 90. They showed that risk aversion increased with these higher stakes and that the RRA of one third of subjects was in the highest RRA range when stakes are increased by a factor of 90. Harrison at al. (2005) use a design similar to Holt and Laury (2002) but control for order effects. They find that stakes effects are only significant between a factor of 20 and 50 (or 90), but not between 50 and 90. Of course, if incentives increase even further beyond a factor of 50,

---

[21]Here, α is the measure of RRA.

[22]Note that this is also the modal switching point under low stakes in Holt and Laury (2002).

[23]Note that in Holt and Laury (2002) there are also less inconsistent subjects under high stakes and more inconsistencies under comparable hypothetical payoffs.

only few subjects need such high incentives to exert their required cognitive effort and we should expect that the increase in risk aversion is not significant anymore. Although this cognitive effort explanation may explain the results of our experiment, we did not design the experiment to test this explanation. Further research should therefore investigate this potential explanation more closely in order to answer the question what the true level of risk aversion is.

5.3   Robustness

One may be concerned that the differences in the RRA estimates between the HL method and our method may be due to a preference for switching in the same row across tables. For instance, if a subject always switches after row 5 (from A to B in the HL tables and from B to A in our tables) we would measure a higher RRA in our tables. When comparing the HL tables to our tables, there are always the same number of rows where identical switching behavior would induce a higher RRA with our method as there are where identical switching leads to a higher RRA with the HL method.[24] Nevertheless, if a majority of subjects switched in exactly those rows where we would measure a higher RRA for our tables, Result 3 may simply be explained by a preference for identical switching behavior.

For all four comparisons of Figure 2, we find that only a minority of subjects switches in the same row in both tables. The respective numbers are 19.7%, 9.9%, 10.6%, and 12.7% for Panels 2A, 2B, 2C, and 2D.[25] Nevertheless, of those subjects that switch in the same row a majority switches in those rows that induce a higher RRA with our method. For the respective Panels 2A, 2B, 2C, and 2D 80%, 85.7%, 71.4%, and 55.6% of the 'same row switchers' switch in rows where our tables yield a higher RRA, and only 13.3%, 0%, 28.6%, and 11.1% switch in rows where the HL tables yield a higher RRA. It may therefore be possible that a preference for identical switching behavior drives Result 3.

However, we can test whether Result 3 still holds for those subjects that switch in different rows. If we exclude the 'same row switchers', we can test whether Result 3 is driven by a preference for switching in the same row across both types of tables. For the comparisons

---

[24]When comparing original HL tables to our adjusted tables (HLol vs. MRal and HLoh vs. MRah), switching after row 1, 5, and 6 yields a higher RRA in our tables, switching after row 3, 4, 8, and 9 yields a higher RRA in the HL tables, and switching after row 2 and 7 yields the same RRA in both tables. Never switching (choosing always B in both tables) also leads to a higher RRA in our tables. When comparing our original to the adjusted HL tables (HLal vs. MRol and HLah vs. MRoh), switching after row 3, 4, and 5 yields a higher RRA in our tables, switching after row 2, 7, 8, and 9 yields a higher RRA in the HL tables, and switching after row 1 and 6 leads to the same RRA in both tables. Again, never switching and choosing always B yields a higher RRA in our tables.

[25]Note that all subjects answered not only two but four tables. And there are only two subjects that switch in the same row in all four tables.

of Figure 2, we again perform two-sided Wilcoxon signed rank sum tests where we exclude 'same row switchers'.[26] For all comparisons we observe that differences are still significant. For the comparison of HLol vs. MRal (Panel 2A) we get $z = -4.540$, $p = 0.0000$, $N = 61$. For the comparison of HLoh vs. MRah (Panel 2B) we get $z = -1.787$, $p = 0.0740$, $N = 64$. The comparison of HLal vs. MRol (Panel 2C) yields $z = -3.908$, $p = 0.0001$, $N = 59$, and the comparison of HLah vs. MRoh (Panel 2D) yields $z = -3.982$, $p = 0.0001$, $N = 62$. Thus, the differences of Panels 2A, 2C, and 2D are still significant at the 1%-level, and the difference of Panel 2B is still significant at the 10%-level. We can now state the following result.

Result 5 *Result 3 is not due to a preference for switching in the same row across both methods.*

Since we had eight different treatments in our experiment, we can also test whether the order in which the tables were presented to the subjects had an effect on the elicited intensity of risk aversion. This could be the case if subjects are primed to think about all tables in terms of the first table they completed. To control for such order effects, we first test for each of the comparisons of Figure 2 whether it made a difference if a specific table was completed *before* or *after* its corresponding 'other-type' table. As an example, consider the HLol table. Here, we test whether the RRA distribution of HLol in treatment 1 and 5 (where it is completed before MRal) is different to the RRA distribution of Hlol in treatment 3 and 7 (where it is completed after MRal). For each of the eight different tables we perform a two-sided Kolmogorov-Smirnov test on the equality of RRA distributions and find no significant difference in either table.[27]

To control for order effects in the comparisons of Figure 3, we also test for each table whether it made a difference if it was completed *before* or *after* its corresponding 'other-stake' table. Consider again as an example the HLol table. Here, we test whether the RRA distribution of HLol in treatment 1 and 3 (where it is completed before HLoh) is different to the RRA distribution of HLol in treatment 5 and 7 (where it is completed after HLoh). Again, none of the two-sided Kolmogorov-Smirnov tests shows a significant difference.[28] We

---

[26]When performing two-sided Kolmogorov-Smirnov tests instead, we observe similar results. All comparisons of Figure 2 stay significant at the 1%-level ($p = 0.000$, $N = 122$ for Panel 2A; $p = 0.008$, $N = 128$ for Panel 2B; $p = 0.000$, $N = 118$ for Panel 2C; and $p = 0.000$, $N = 124$ for Panel 2D).

[27]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.199$, $N = 107$), HLal ($p = 0.822$, $N = 94$), HLoh ($p = 1.000$, $N = 103$), HLah ($p = 0.949$, $N = 95$), MRol ($p = 0.424$, $N = 78$), MRal ($p = 0.896$, $N = 79$), MRoh ($p = 0.545$, $N = 73$), and MRah ($p = 0.763$, $N = 76$).

[28]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.188$, $N = 107$), HLal ($p = 0.949$, $N = 94$), HLoh ($p = 0.990$, $N = 103$), HLah ($p = 0.940$, $N = 95$), MRol ($p = 0.796$, $N = 78$), MRal ($p = 0.979$, $N = 79$), MRoh ($p = 0.775$, $N = 73$), and MRah ($p = 0.957$, $N = 76$).

are now able to summarize these results on order effects. The order effects that are captured by Result 6 test whether high-stakes tables are completed before or after corresponding low-stakes tables (and vice versa) as well as whether original (HL and MR) tables are completed before or after corresponding adjusted (MR and HL) tables (and vice versa).

Result 6 *The order in which tables were completed by subjects does not influence the elicited intensities of risk aversion.*

The analysis of sections 5.1 and 5.2 excluded subjects that switched multiple times from one option to another, or chose the first-order dominated option, and thus were inconsistent in completing a table. Over all eight different tables, we find that on average 24% of subjects are inconsistent which is in line with other studies.[29] However, Holt and Laury (2002) treated those inconsistent subjects differently than we did. They simply counted how often subjects chose each option (even if the option was not only chosen in subsequent rows) and assumed the inconsistency away. If we treat our inconsistent subjects in the same way, we can test whether the inconsistent subjects show a different pattern of risk aversion than the consistent subjects. When performing Kolmogorov-Smirnov tests, we do not find any significant difference in the distributions of consistent and inconsistent subjects.[30]

6 Conclusion

The multiple price-list method of HL has become *the* standard way to measure the intensity of individuals' risk attitudes in experiments. Several other methods which also involve choices between lotteries have been proposed to accomplish a similar task. Binswanger (1980, 1981) asks subjects to choose between pairs of different 50/50 gambles, where higher

---

[29]Of those 24%, 18% switched multiple times and 6% chose the first-order dominated option. Bruner et al. (2008), for instance, find 30% inconsistent subjects where 25% are multiple switchers and 5% have non-increasing utility. As noted by Andersen et al. (2006, p386) "it is quite possible that [multiple] switching behavior is the result of the subject being indifferent between the options. The implication here is that one simply use a "fatter" interval to represent this subject in the data analysis, defined by the first row that the subject switched at and the last row that the subject switched at." When allowing for indifference, Andersen et al. (2006) find that multiple switching is only 5.8% and 24.3% choose the indifference option. When imposing a single switching point, choosing indifference increases to 30%, which again indicates that multiple switching behavior is caused by 'thicker' indifference curves. In contrast, Bruner (2007) suggests an instructional variation emphasizing that only *one* decision will determine earnings (and further emphasizing incentive compatibility of the payment rule) and finds multiple switching reduced, from 25.8% to 6.7% and from 13.3% to 2.3% in two different elicitation formats. This would rather suggest that multiple switching is caused by errors.

[30]The results of the two-sided Kolmogorov-Smirnov tests are: HLol ($p = 0.916$, $N = 116$), HLal ($p = 0.390$, $N = 113$), HLoh ($p = 0.328$, $N = 116$), and HLah ($p = 0.180$, $N = 116$). Subjects were only excluded in case they never chose option B (3 subjects in HLal), since then even reshuffling their choices cannot make them consistent.

expected values come at the cost of higher standard deviations. Eckel and Grossman (2008) use a gamble design similar to Binswanger (1980, 1981) but subjects choose one out of five different 50/50 gambles. Hey and Orme (1994) ask a battery of 100 choices with more complex probabilities. None of these approaches uses mean preserving spreads *among* different choice situations in a way similar to our approach. Since it is exactly this feature that allows us to classify subjects as more or less risk-averse independently of a specific model, other elicitation methods need to impose the same utility structure on all individuals in order to classify them. None of the methods is model-independent in that it can rank individuals as more or less risk-averse when they have heterogeneous risk preferences, such as EUT *and* non-EUT. This seems especially problematic in light of existing evidence.

After modifying the HL method and thereby proposing a new model-independent multiple price-list method to elicit the intensity of individuals' risk attitudes, we further compared our proposed method to the HL method in a lab experiment. Our results for the HL method replicate the findings of Holt and Laury (2002). Furthermore, concerning the direction of risk attitude (i.e. risk-averse, risk-neutral, or risk-seeking) we find that individuals are classified the same across both methods.

However, with our method we found systematic differences concerning the intensity of risk attitude. The classification of individuals as being more or less risk-averse than other individuals is quite different across methods. This is important for experimental studies that use the HL method (and thereby assume EUT for the classification) in order to control for risk aversion in observed behavior. Concerning the level of individuals' intensity of risk aversion we found that our method yields higher measures of risk aversion that are much closer to what is observed in the field. Furthermore, while increasing the stakes increases risk aversion with the HL method, our method is robust toward such stakes effects. We also offered a cognitive effort explanation of our results that needs to be tested in future research. This may help to answer the important question of what subjects' true level of risk aversion is after all.

Appendix A: The General Model-Independent Method

Define row i as consisting of two alternatives A_i and B_i. Each alternative consists of two possible outcomes, $A1_i$ and $A2_i$ of alternative A_i and $B1_i$ and $B2_i$ of alternative B_i. Given row i is the row played (if only one row of the table is randomly selected to be played), each outcome of each alternative is realized with probability $1/2$. Choose a row $i = m$ and define $A1_m \equiv a, A2_m \equiv b, B1_m \equiv c, B2_m \equiv d$, where $a, b, c, d \in \mathbb{R}$. Without loss of generality, choose a and b such that $a < b$, and c and d such that $c < d$.[31] Our method of elicitation requires that $a \leq c$, $b \leq d$, and at least one relation holds strictly. It follows that

$$A_m \prec_{FSD} B_m. \tag{1}$$

(1) means that B_m first-order stochastically dominates (FSD) A_m.[32] Every individual with strictly increasing utility prefers B_m over A_m.

Then define values of row $m + 1$ as follows: $A1_{m+1} \equiv a, A2_{m+1} \equiv b, B1_{m+1} \equiv c - k_1, B2_{m+1} \equiv d + k_1$. Further define values of row $m + 2$ in the following way: $A1_{m+2} \equiv a, A2_{m+2} \equiv b, B1_{m+2} \equiv c - k_1 - k_2, B2_{m+2} \equiv d + k_1 + k_2$. Continue these mean preserving spreads (MPS's) of option B until the last row is reached. Generally, for $\check{n} > 0$ we can define $A1_{m+\check{n}} \equiv a, A2_{m+\check{n}} \equiv b, B1_{m+\check{n}} \equiv c - k_1 - k_2 - \ldots - k_{\check{n}}, B2_{m+\check{n}} \equiv d + k_1 + k_2 + \ldots + k_{\check{n}}$, where $k_1, k_2, \ldots \in \mathbb{R}^+$.

In a similar way, we can define values of row $m - 1$ as follows: $A1_{m-1} \equiv a - k^1, A2_{m-1} \equiv b + k^1, B1_{m-1} \equiv c, B2_{m-1} \equiv d$. Now define values of row $m - 2$ as follows: $A1_{m-2} \equiv a - k^1 - k^2, A2_{m-2} \equiv b + k^1 + k^2, B1_{m-2} \equiv c, B2_{m-2} \equiv d$. Again, these MPS's of option A can be continued until the first row is reached. Generally, for $\hat{n} > 0$ define $A1_{m-\hat{n}} \equiv a - k^1 - k^2 - \ldots - k^{\hat{n}}, A2_{m-\hat{n}} \equiv b + k^1 + k^2 + \ldots + k^{\hat{n}}, B1_{m-\hat{n}} \equiv c, B2_{m-\hat{n}} \equiv d$, where $k^1, k^2, \ldots \in \mathbb{R}^+$.

With these definitions it follows that for all i the expected values are $\mathbb{E}[A_i] = \frac{a+b}{2}$ and $\mathbb{E}[B_i] = \frac{c+d}{2}$. Table A illustrates our generalized elicitation method.

From $k_1, k_2, \ldots > 0$ and $k^1, k^2, \ldots > 0$ it follows that for all $i \geq m$ we can state $B1_i > B1_{i+1}$ and $B2_i < B2_{i+1}$. Also, for all $i \leq m$ it holds that $A1_i > A1_{i-1}$ and $A2_i < A2_{i-1}$.

Let us first consider all rows with $i \geq m$. Since all $A1_i$ and $A2_i$ are identical, it follows

---

[31]We chose to make these inequalities strict in order to avoid 'certainty effect' issues.
[32]In order to make this more salient, we even used state-wise dominance as a special case of FSD in the experiment.

Table A: Our Generalized Elicitation Method

Row No. i	Option A_i		Option B_i	
	Prob. 1/2 Outcome $A1_i$	Prob. 1/2 Outcome $A2_i$	Prob. 1/2 Outcome $B1_i$	Prob. 1/2 Outcome $B2_i$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$m-\hat{n}$	$a-k^1-k^2-\ldots-k^{\hat{n}}$	$b+k^1+k^2+\ldots+k^{\hat{n}}$	c	d
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$m-2$	$a-k^1-k^2$	$b+k^1+k^2$	c	d
$m-1$	$a-k^1$	$b+k^1$	c	d
m	a	b	c	d
$m+1$	a	b	$c-k_1$	$d+k_1$
$m+2$	a	b	$c-k_1-k_2$	$d+k_1+k_2$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$m+\check{n}$	a	b	$c-k_1-k_2-\ldots-k_{\check{n}}$	$d+k_1+k_2+\ldots+k_{\check{n}}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

that

$$A_m = A_{m+1} = \ldots = A_{m+\check{n}}$$

$$\Rightarrow \quad A_m \sim A_{m+1} \sim \ldots \sim A_{m+\check{n}}. \tag{2}$$

Since $B1_{m+\check{n}} = B1_{m+\check{n}-1} - k_{\check{n}}$ and $B2_{m+\check{n}} = B2_{m+\check{n}-1} + k_{\check{n}}$ it follows that $B_{m+\check{n}}$ is a MPS of $B_{m+\check{n}-1}$.[33] We use the standard (behavioral) definition of risk aversion and say that an individual is risk-averse if she dislikes increases in risk, i.e. if she dislikes MPS's. We thus employ the increasing risk definitions of Rothschild and Stiglitz (1970). For every risk-averse individual

$$B_m \prec_{MPS} B_{m+1} \prec_{MPS} \ldots \prec_{MPS} B_{m+\check{n}}$$

$$\Rightarrow \quad B_m \succ B_{m+1} \succ \ldots \succ B_{m+\check{n}}. \tag{3}$$

Combining (1), (2), and (3) we can derive that the choice between $A_{m+\check{n}}$ and $B_{m+\check{n}}$ for every $\check{n} > 0$ is a trade-off between the advantage of the FSD-improvement from B_m over A_m ($= A_{m+\check{n}}$) and the disadvantage of the increase(s) in risk (MPS('s)) when going from B_m to $B_{m+\check{n}}$.[34]

Suppose an individual chooses $A_{m+\check{n}}$ over $B_{m+\check{n}}$, but chooses $B_{m+\check{n}-1}$ over $A_{m+\check{n}-1}$. Then, the MPS's from row m until row $m+\check{n}-1$ were not enough to distract him from the FSD-improvement of B_m over A_m. In contrast, the MPS's from row m until row $m+\check{n}$ were enough

---

[33]In going from $B_{m+\check{n}-1}$ to $B_{m+\check{n}}$, general MPS's could be applied. However, since we defined each option as having two possible outcomes only (where each occurs with probability 1/2) a MPS can only be attained by adding and subtracting k_n. In giving up that all outcomes of the table are equally likely, one would not only require that subjects understand situations other than certainty and equally likely, but one would also introduce complications such as probability weighting.

[34]Note that one could also use mean preserving contractions (i.e. decreases in risk) when going from A_m to $A_{m+\check{n}}$, either instead or in addition to the MPS's that are used when going from B_m to $B_{m+\check{n}}$. This would be useful if one wanted to induce similar changes between the two alternatives over the rows of the table.

to outweigh the FSD-improvement of B_m over A_m. So, we can define an 'intermediate' hypothetical option $\bar{B}_{m+\check{n}}$ with $\bar{B}1_{m+\check{n}} = B1_{m+\check{n}-1} - \kappa k_{\check{n}}$ and $\bar{B}2_{m+\check{n}} = B2_{m+\check{n}-1} + \kappa k_{\check{n}}$. The $\kappa \in [0;1]$ is chosen such that $A_{m+\check{n}} \sim \bar{B}_{m+\check{n}}$. Put differently, κ is the fraction of $k_{\check{n}}$ that would make the individual indifferent between $A_{m+\check{n}}$ and $B_{m+\check{n}}$ if $\kappa k_{\check{n}}$ instead of $k_{\check{n}}$ would have been used. For this individual $\bar{B}_{m+\check{n}}$ is then a mean utility preserving spread of $A_{m+\check{n}}$ in the sense of Diamond and Stiglitz (1974).

Now, suppose there is another individual who is less risk-averse. This individual would then still strictly prefer $\bar{B}_{m+\check{n}}$ over $A_{m+\check{n}}$. It follows, that this individual would choose at least as many times B over A as the more risk-averse individual.[35] The smaller the k_i's, the finer are the differences in risk aversion that can be observed. Any observed differences in behavior must be due to (sufficiently strong) differences in risk aversion.

In order to distinguish risk-seeking individuals, consider all rows with $i \leq m$. Since all $B1_i$ and $B2_i$ are identical, it follows that

$$B_m = B_{m-1} = \ldots = B_{m-\hat{n}}$$
$$\Rightarrow \quad B_m \sim B_{m-1} \sim \ldots \sim B_{m-\hat{n}}. \tag{4}$$

Since $A1_{m-\hat{n}} = A1_{m-\hat{n}+1} - k^{\hat{n}}$ and $A2_{m-\hat{n}} = A2_{m-\hat{n}+1} + k^{\hat{n}}$ it follows that $A_{m-\hat{n}}$ is a MPS of $A_{m-\hat{n}+1}$. Of course, every risk-seeking individual likes MPS's and thus

$$A_m \prec_{MPS} A_{m+1} \prec_{MPS} \ldots \prec_{MPS} A_{m-\hat{n}}$$
$$\Rightarrow \quad A_m \prec A_{m-1} \prec \ldots \prec A_{m-\hat{n}}. \tag{5}$$

Combining (1), (4) and (5) we can derive that the choice between $A_{m-\hat{n}}$ and $B_{m-\hat{n}}$ for every $\hat{n} > 0$ is a trade-off between the advantage of the FSD-improvement from B_m ($= B_{m-\hat{n}}$) over A_m and the advantage of the increase(s) in risk when going from A_m to $A_{m-\hat{n}}$. By similar arguments as before, it follows that a less risk-seeking individual would choose at least as many times B over A as the more risk-seeking individual.

As in the HL tables, for risk-averse as well as for risk-seeking individuals it is optimal to switch options only once. Risk-averse individuals switch from B to A after row m (the earlier the more risk-averse) and risk-seeking individuals switch from A to B before row m (the earlier the less risk-seeking).[36] For risk-neutral individuals it is optimal to choose option B throughout since it offers the higher expected payoff.

---

[35]Strictly speaking, an individual who chooses as many times B over A as another individual could still be slightly more or slightly less risk-averse. In this case her κ would be lower or higher, respectively.

[36]Of course, if one wanted to induce switching from A to B for most (i.e. risk-averse) subjects as in the HL tables, one could achieve this by just exchanging alternatives A and B.

References

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström (2006). "Elicitation Using Multiple Price List Formats", *Experimental Economics* 9(4), pp383-405.

Aumann, Robert J., and Roberto Serrano (2008). "An Economic Index of Riskiness", *Journal of Political Economy* 116(5), pp810-836.

Binswanger, Hans P. (1980). "Attitudes toward Risk: Experimental Measurement In Rural India", *American Journal of Agricultural Economics* 62(3), pp395-407.

Binswanger, Hans P. (1981). "Attitudes toward Risk: Theoretical Implications of an Experiment in Rural India", *Economic Journal* 91(4), pp867-890.

Bombardini, Matilde, and Francesco Trebbi (2010). "Risk Aversion and Expected Utility Theory: An Experiment with Large and Small Stakes", Working Paper, University of British Columbia.

Bruner, David (2007). "Multiple Switching Behavior in Multiple Price Lists", Working Paper, Appalachian State University.

Bruner, David, Michael McKee, and Rudy Santore (2008). "Hand in the Cookie Jar: An Experimental Investigation of Equity-Based Compensation and Managerial Fraud", *Southern Economic Journal* 75(1), pp261-278.

Brunnermeier, Markus K., and Stefan Nagel (2008). "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation", *American Economic Review* 98(3), pp713-736.

Calvet, Laurent E., and Paolo Sodini (2010). "Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios", NBER Working Paper, No. w15859.

Chetty, Raj (2006). "A New Method of Estimating Risk Aversion", *American Economic Review* 96(5), pp1821-1834.

Chew, Soo Hong, Edi Karni, and Zvi Safra (1987). "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities", *Journal of Economic Theory* 42(2), pp370-381.

Diamond, Peter A., and Joseph E. Stiglitz (1974). "Increases in Risk and in Risk Aversion", *Journal of Economic Theory* 8(3), pp337-360.

Eckel, Catherine C., and Philip J. Grossman (2008). "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices", *Journal of Economic Behavior and Organization* 68(1), pp1-17.

Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert (2010). "Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size", *Journal of Risk and Uncertainty* 40(2), pp147-180.

Fischbacher, Urs (2007). "Z-Tree: Zurich Toolbox for Ready-Made Economics Experiments", *Experimental Economics* 10(2), pp171-178.

Friend, Irwin, and Marshall E. Blume (1975). "The Demand for Risky Assets", *American Economic Review* 65(5), pp900-922.

Greiner, Ben (2004). "An Online Recruitment System for Economic Experiments", in Kurt Kremer, and Volker Macho, eds., *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, pp79-93.

Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutsröm (2005). "Risk Aversion and Incentive Effects: Comment", *American Economic Review* 95(3), pp900-904.

Harrison, Glenn W., Steven J. Humphrey, and Arjan Verschoor (2010). "Choice under Uncertainty: Evidence from Ethiopia, India and Uganda", *Economic Journal* 120(2), pp80-104.

Hey, John D., and Chris Orme (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data", *Econometrica* 62(6), pp1291-1326.

Hildreth, Clifford, and Glenn J. Knowles (1982). "Some Estimates of Farmers' Utility Functions", *Technical Bulletin* 335, Agricultural Experimental Station, University of Minnesota, Minneapolis.

Holt, Charles A., and Susan K. Laury (2002). "Risk Aversion and Incentive Effects", *American Economic Review* 92(5), pp1644-1655.

Kydland, Finn E., and Edward C. Prescott (1982). "Time to Build and Aggregate Fluctuations", *Econometrica* 50(6), pp1345-1370.

Levy, Haim (1994). "Absolute and Relative Risk Aversion: An Experimental Study", *Journal of Risk and Uncertainty* 8(3), pp289-307.

Machina, Mark J. (1982). ""Expected Utility" Analysis Without the Independence Axiom", *Econometrica* 50(2), pp277-323.

Machina, Mark J. (2008). "Non-Expected Utility Theory", in Steven N. Durlauf, and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., vol. 6, New York: Palgrave MacMillan, pp74-84.

McCord, Mark, and Richard de Neufville (1986). ""Lottery Equivalents": Reduction of the Certainty Effect Problem in Utility Assessment", *Management Science* 32(1), pp56-60.

Mehra, Rajnish (2003). "The Equity Premium: Why Is It a Puzzle?", *Financial Analysts Journal* 59(1), pp54-69.

Quiggin, John (1982). "A Theory of Anticipated Utility", *Journal of Economic Behavior and Organization* 3(4), pp323-343.

Rothschild, Michael, and Joseph E. Stiglitz (1970). "Increasing Risk: I. A Definition", *Journal of Economic Theory* 2(3), pp225-243.

Schmidt, Ulrich, and Horst Zank (2008). "Risk Aversion in Cumulative Prospect Theory", *Management Science* 54(1), pp208-216.

Szpiro, George (1986). "Measuring Risk Aversion: An Alternative Approach", *Review of Economics and Statistics* 68(1), pp156-159.

Tobin, James, and Walter Dolde (1971). "Wealth, Liquidity and Consumption", in *Consumer Spending and Monetary Policy: The Linkage*, Boston: Federal Reserve Bank of Boston, pp99-146.

Tversky, Amos, and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty* 5(4), pp297-323.