Measuring Normative Risk Preferences*

Gosse A.G. Alserda[1]

[1]Erasmus School of Economics, Erasmus University Rotterdam

February 7, 2017

Abstract

The results of eliciting risk preferences depend on the elicitation method. Different methods of measuring the same variable tend to produce different results. This raises the question whether normative risk preferences can be elicited at all. Using two types of manipulation, I assess the normative value of risk preference elicitation methods. Following IRT, the results of the multiple lottery choice method are combined with two qualitative methods into a composite score. The responses of 9,235 pension fund members to a dedicated survey indicate this composite score approximates the latent variable normative risk preferences better than individual method responses do, substantially reducing measurement noise and method-specific biases. Analysis of the manipulations shows that both the results and the normative value of the risk preference elicitation methods depend on the specific amounts, order, and endowment chosen. Combining simpler methods with more advanced methods framed closely to the relevant situation increases the normative value of elicited risk preferences.

Keywords: Normative Risk Preferences, Composite Score, Multiple Lottery Choice, Item Response Theory, Manipulations

Words: 7,963

*I thank Benedict Dellaert, Rogier Potter van Loon, and Fieke van der Lecq for valuable suggestions and the participants at the ESA World Meeting 2016 for their comments. I am grateful to Korn Ferry Hay Group NL and several employers for their fruitful cooperation on this project.

1 Introduction

Risk preferences are relevant to a large number of decisions, including financial decisions, which often involve trading off risk and return. Research has shown that risk preferences tend to differ between individuals (Harrison et al., 2007; Holt and Laury, 2002; Weber et al., 2002) and that differences in risk preferences affect optimal choices (Campbell et al., 2003; Viceira, 2001). Risk preferences can vary between persons for a number of reasons, including parents' risk taking (Levin and Hart, 2003), genetic variation (Zyphur et al., 2009), nationality (Hsee and Weber, 1999), age (Yao et al., 2011), and gender (Jianakoplos and Bernasek, 1998).

Individuals often make choices without explicitly knowing (the quantitative value of) their risk preferences. However, in the case of delegated decisions, risk preferences need to have explicit value for the delegated decision makers to make decisions that maximize value for the relevant person(s). Since most individuals are unaware of their risk preferences, these must be elicited indirectly. Many risk preference elicitation methods are cited in the literature, including the multiple lottery choice (MLC) method (Holt and Laury, 2002), the betting game choice method (Gneezy and Potters, 1997), and the willingness-to-pay (Becker et al., 1964) and self-description methods (e.g. Kapteyn and Teppa (2011)). However, different risk preference elicitation methods tend to yield distinct results for the same variable and within the same domain (Alserda et al., 2017). Even the same method can provide different results due to framing (Harrison et al., 2007; Lévy-Garboua et al., 2012), the domain of the question (Weber et al., 2002), or noisy behaviour (Dave et al., 2010).

This raises the question whether it is possible to measure normative risk preferences and, if so, how. Normative preferences are described as 'preferences that represent an economic agent's true interests' (Beshears et al., 2008, p. 3), as opposed to revealed preferences, which are 'preferences that rationalize an economic agent's observed actions' (p. 2). In the case of measurement noise or biases in the elicitation of risk preferences, revealed preferences can differ from normative preferences. The extent to which revealed risk preferences correspond to normative risk preferences indicates the method's normative value. However, normative risk preferences cannot be measured; therefore, normative risk preferences are a latent variable and need to be approximated using revealed preferences.

Delegated decision makers are increasingly expected to ensure that decisions reflect individual preferences (EIOPA, 2013; Frijns, 2010; Rozinka and Tapia, 2007). Delegated decision makers should thus become familiar with the normative preferences. The use of revealed risk preferences that are not in line with normative preferences will lead to suboptimal decisions, which will lower individuals' welfare (Viceira, 2001).

A fundamental trade-off within risk preference elicitation methods is that of simplicity versus practical usefulness. More complicated methods involving monetary amounts and probabilities have better predictive accuracy (Dave et al., 2010) and can be easily translated to relative risk aversion (RRA), a quantitative measure of risk aversion. Preferences can thus be easily applied in practical situations, as in setting an asset allocation (Viceira, 2001). However, more complicated methods can also induce noisier behaviour, especially for subjects with lower numeracy skills, who may not fully understand their options (Dave et al., 2010). The normative value of more complicated methods may therefore be lower.

The simplest methods include self-description questions. In these methods, respondents are asked to describe themselves, normally in comparison to others. The respondents do not need to be financially literate to understand such questions; however, the expectations of others may not be constant and true risk preferences may not be known. The results are thus hard to quantify and therefore difficult to apply to real-life situations, as in asset allocation (Alserda et al., 2017).

Van Rooij et al. (2011) find that large proportions of the population have only a basic understanding of financial topics. Lusardi and Mitchell (2007) confirm this finding and show that only half of older Americans can correctly answer questions about compounding, inflation, and risk diversification. Financial literacy influences financial decision making (Van Rooij et al., 2011) and financially illiterate individuals have difficulty making adequate retirement plans (Lusardi and Mitchell, 2007). Intuitively, respondents who have difficulty understanding basic financial topics cannot be expected to perfectly understand complicated risk elicitation questions. Therefore, one should be careful interpreting the results of more complicated risk elicitation methods.

Beshears et al. (2008) confirm this intuition and show that complexity increases the effect of biases in the elicitation of risk preferences. If the number of investment options increases, experimental subjects are shown to be more likely to choose simple, risk-free investment options compared to complex, risky investment options (Iyengar and Kamenica, 2006). Greater complexity in the form of more options also tends to reduce pension plan participation (Beshears et al., 2013; Iyengar et al., 2004).

In this paper, I elicit risk preferences with three different methods, with two variations of the last method: an augmented version of the MLC method, an investment choice question, and two self-description questions. Using item response theory (IRT), the results of the three methods are combined into a single composite risk aversion score. This composite score is the closest available approximation to the latent variable normative risk preferences (Menkhoff and Sakha, 2016) and can benefit from both the simplicity of the self-description method and the quantitative value of the MLC method.

Two kinds of manipulation are added to the risk preference elicitation methods. First, the MLC method is manipulated in terms of order, amounts, and starting probability. The composite score is used as a reference point to compare the normative value of the four manipulations of the MLC method. Second, the subjects are split into a group where the (risk-free) base pension is included in the question and a group where it is excluded. Both manipulations indicate the extent to which framing effects influence the normative value of elicitation methods.

Risk preferences are elicited in the pension domain. The pension domain involves a prominent case of delegated decision making, with an investment manager allocating the assets of pension fund participants (e.g. pension capital). The optimal mix of assets depends strongly on risk preferences, which must therefore be known to optimize the asset mix (Campbell et al., 2003; Viceira, 2001).

Elicitation of the risk preferences of 9,235 pension fund participants confirms that different elicitation methods result in different elicited risk preferences. Combining multiple risk preference elicitation methods reduces measurement noise and method-specific biases. The composite score of risk preferences therefore provides a more reliable estimation of normative risk preferences. Of the individual methods, the augmented MLC method, especially with a manipulated starting point, provides the most useful quantitative information in the domain of pension income. The absence of a strong effect of the inclusion or exclusion of the base pension shows that it is important to elicit risk preferences in a situation (*i.e.*, endowment) as closely as possible to the observed reality for members, since individuals have difficulty processing these kinds of effects. Risk preferences are dependent on sociodemographic information, such as income, age, and education level.

2 Method

Risk preferences are elicited with three different methods and a number of variations on these methods. The methods include the MLC method (Hardy, 2001), the self-description method (Kapteyn and Teppa, 2011), and the investment choice method (Van Rooij et al., 2007).

2.1 MLC

In line with previous research (Anderson and Mellor, 2009; Croson and Gneezy, 2009; Dohmen et al., 2011), the MLC method of Holt and Laury (2002) is used as the primary method for eliciting risk preferences in the pension domain. The MLC method presents respondents a series of choices between two lotteries. The first lottery is safe and has a smaller difference between the good state and the bad state. The second lottery is riskier and has a larger difference between both states. At the start, the probability of the good state is low and the safe lottery is dominant for all but extremely risk-seeking individuals. In subsequent choices, the probability of the good state increases and the risky lottery becomes increasingly attractive. At a certain point, the respondents will switch from the safe lottery to the risky lottery, the switching point thereby revealing their risk preferences.

However, I implement a number of improvements to cope with much observed irregularities. First, in line with Weber et al. (2002), the questions are adjusted to the relevant domain (i.e. pensions). Second, the range of possible RRA values is increased to deal with the higher risk aversion expected in the pension domain (Van Rooij et al., 2007). To keep the maximum number of questions constant and to reduce the predictability of the subsequent questions, the presented possibilities of the good state are changed compared to those of Holt and Laury (2002) (see Table 1). To prevent respondents from changing between the safe and risky lottery more than once, the online survey is programmed so that when respondents change from the safe lottery to the risky lottery, they must confirm their choice. After confirmation, their switching point is recorded and they proceed to the next survey question. Respondents can return to a question and change their choice until they have confirmed it. An example of such a question is presented in Figure 1.

In addition, prior to the effective question, respondents must answer an introductory MLC question. This question introduces the respondents to the concept of this method and allows them to make three choices between two lotteries concerning holiday trips. The results of this question are not used but should increase understanding of the question and reduce previously observed biases. (Holt and Laury, 2002; Harrison et al., 2007).

Figure 1: Example of an adjusted MLC question

You have indicated that your after-tax monthly income is between €1,800 and €2,000. The amounts in this question are based on this income level. They represent monthly net income levels, including the state old age pension. The probabilities changes with your choice.

At which probability would you switch from Plan A to Plan B?

	Plan A	**Plan B**
Your guaranteed income is:	$1,290	$860
In addition, you have a probability of of receiving additional income of:	**10%** $220	**10%** $1,080
So you have a **10%** probability of a total pension income of:	$1,510	$1,940

Plan A **Plan B**

Notes: Example for a participant with a net monthly income of $2,150. This example represents the first choice out of a sequence of 10 in which the probability (highlighted in red) systematically increases for the additional pension income. All amounts are converted to US dollars.

2.2 Manipulations

To assess the effect of theoretically neutral variations in the presentation of the MLC questions, two kinds of manipulation are added. First, 70% of the population are presented the baseline question, using the numbers discussed above, but 10% are instead presented a version with improved amounts for the risky lottery. The risky lottery now has probabilities of 50% and 100% in the bad and good states, respectively, increasing both the minimum and maximum discernible RRA. Another 10% are presented the same amounts as in the baseline scenario, but with the opposite presentation direction, that is, these respondents start with a dominant risky lottery choice and can reveal their risk preferences by switching towards the safe lottery. The last 10% of respondents are presented the baseline scenario with a different starting value. The first question in this manipulation has a 50%/50% probability for the good and bad states.

Manipulation 1: Baseline question. The safe lottery provides 60% or 70% of one's income and the risky lottery 40% or 90%. It starts with a 10% probability for the good state and this probability increases with each subsequent question.

Manipulation 2: The risky lottery is changed to a 50% probability in the bad state and 100% in the good state, so the range of measurable RRA values increases upwards.

Manipulation 3: The series follows the opposite direction, respondents start with a dominant risky lottery (i.e. 100% probability for the good state), and the probability for the good state decreases with subsequent questions.

Manipulation 4: The lottery starts with a 50% probability for the good state, which increases again with subsequent questions.

The second type of manipulation concerns the inclusion of the state's base pension. Half of the respondents are presented the questions with the amounts representing full pensions (occupational pensions and base pension). The other half is presented the same amounts, but for occupational pensions only. Since the base pension is between \$862 and \$1,696 a month, this would, rationally, lead to substantial differences in the revealed risk preferences. This manipulation is only implemented for pension funds 1 and 3.

RRA

The points at which respondents switch between the risky and the safe lotteries can be translated into a range of RRA levels. This range is obtained by calculating the RRA that makes the respondent indifferent to the trade-off switching point (first choice for the risky lottery) and the RRA that makes the respondent indifferent to the question before that (last choice for the safe lottery) (Holt and Laury, 2002). This approach yields the RRA ranges presented in Table 1, which depend on the manipulation. To obtain the distribution of RRA, a uniform distribution is used for the closed range switching points. For the open ranges, without an upper or lower limit, a normal distribution is assumed and observations are distributed in the tail.

Table 1: **Risk aversion scores based on lottery choices**

Switching point (p =)		**RRA Range for** $U(X) = x^{1-r}/(1-r)$		
	Sample 1 *Standard*	**Sample 2** *Alternate values*	**Sample 3** *Opposite direction*	**Sample 4** *Different starting value*
10%	$r < -4.82$	$r < -2.49$	$r < -4.82$	
30%	$-4.82 < r < -1.82$	$-2.49 < r < 1.00$	$-4.82 < r < -1.82$	
50%	$-1.82 < r < 0.00$	$1.00 < r < 2.61$	$-1.82 < r < 0.00$	$r < 0.00$
65%	$0.00 < r < 1.00$	$2.61 < r < 4.13$	$0.00 < r < 1.00$	$0.00 < r < 1.00$
80%	$1.00 < r < 2.85$	$4.13 < r < 6.09$	$1.00 < r < 2.85$	$1.00 < r < 2.85$
90%	$2.85 < r < 4.46$	$6.09 < r < 8.22$	$2.85 < r < 4.46$	$2.85 < r < 4.46$
95%	$4.46 < r < 5.91$	$8.22 < r < 10.22$	$4.46 < r < 5.91$	$4.46 < r < 5.91$
99%	$5.91 < r < 9.03$	$10.22 < r < 14.88$	$5.91 < r < 9.03$	$5.91 < r < 9.03$
100%	$9.03 < r$	$14.88 < r$	$9.03 < r$	$9.03 < r$

Notes: The RRA ranges for different switching points in the MLC question, given one of four samples.

2.3 Self-description questions

Two self-description questions are added to the survey, in line with and based on the work of Kapteyn and Teppa (2011). The first question, resulting in the variable *Careful*, is 'Does the following statement apply to you? My friends describe me as careful'. The answer is on a seven-point Likert scale ranging from totally disagree (1) to totally agree (7). The second question is framed in the pension domain and asks, 'Are you willing to take a risk with your pension contributions?' Answers to this question range from totally agree (1) to totally disagree (7) and results in the variable *Stated aversion.*

Careful (0–7)

My friends describe me as a careful person.

Stated aversion (0–7)

Are you willing to take a risk with your pension?

2.4 Investment choice question

For this question, respondents are required to make a simplified allocation of their fictitious pension capital. For this method, the respondents can allocate their pension allocation over fixed income, described as a fixed return of 2%, and equity, described as risky, with an expected return of 6%. The minimum incremental step is 10%. Although many respondents will find it hard to allocate their assets in line with their normative risk preferences, the question does show how individuals respond to the risk and return trade-off (Lusardi and Mitchell, 2007). Therefore, this measure is expected to add information about normative risk preferences (Van Rooij et al., 2007).

Allocation to bonds (0–100%)

How would you invest your pension contributions?

2.5 Normative risk preferences

Revealed preferences do not necessarily represent normative preferences. Normative preferences are preferences that reflect true interests, whereas revealed preferences rationalize observed behaviour (Beshears et al., 2008). Revealed preferences can differ from normative preferences for several reasons, including individuals having large areas of perceived indifference (Anderson and Mellor, 2009), choices depending on (unstable) emotions Soane and Chmiel (2005), lack of attention (March and Shapira, 1992), or framing effects (LeBoeuf and Shafir, 2003). Different elicitation methods therefore tend to produce different results. Even the same method can produce different risk preference results if applied multiple times (Fellner and Maciejovsky, 2007). This makes the measurement of normative risk preferences far from straightforward.

I use two criteria to assess the normative value of different risk preference elicitation methods and combine information from these methods to create a single measure of risk preference. These criteria are selected because they represent the two reasons for deviations between normative and revealed risk preferences: noise and biases. The following criteria are used to assess the validity of risk preference elicitation methods, specifically in the manipulations of the MLC method.

Criteria:

1. Elicitation methods should have a high correlation with the underlying latent variable, in this case pension domain risk preferences.

2. Elicitation methods should not be sensitive to the (in)ability of respondents to understand the question. Therefore, there should be no strong correlation with financial literacy/education while controlling for other relevant factors, such as human capital.

Unfortunately, the first criterion is difficult to assess, since latent variables are, by definition, not directly observable. The latent variable risk preferences are estimated from the available risk preference elicitation results using principal component factor analysis (CFA) with varimax rotation (Kapteyn and Teppa, 2011). Higher correlations with the factor indicate observed risk preferences that are closer to the normative risk preferences.

In addition, by retaining a second factor and including the level of education in the analysis, the latent variable (financial) literacy, or understanding of the question, is estimated besides the latent variable risk preferences. Elicitation methods that have a higher correlation with this factor are less reliable, since more respondents are likely to not fully understand the question. For observed risk preferences to be in line with normative risk preferences, an elicitation method should have a low correlation with the literacy factor and a high correlation with the risk aversion factor.

A second method to assess the normative value of elicitation methods is by applying IRT. The advantage of IRT is that it allows for a joint estimation of the different samples, since it can cope with missing values. Again, all available risk preference elicitation methods are included, including the different MLC method manipulations. The two relevant outputs from IRT are the discrimination and difficulty coefficients.

Incorporating the information

Using the information of different elicitation methods gives richer insight into the risk preferences of pension fund participants, since measurement noise and method-specific biases are reduced. However, to make the information useful, the results of different elicitation methods should be combined into a single measure of risk preferences.

First, a selection needs to be made for the relevant variables to be included in the composite risk preferences score. Principal component factor analysis with varimax rotation is used for this purpose. Variables are included in the composite score if they have a loading higher than 0.40 (Kapteyn and Teppa, 2011). The factor analysis analyses the joint correlation of separate variables. If these variables describe the same latent variable, but with noise, the noise is (partially) eliminated and the joint correlation will better describe the latent variable.

Second, IRT is used to retrieve latent variable risk aversion from the different methods. In particular, a graded response model is implemented, to connect to the multiple options in each elicitation method that reveal increasing levels of risk aversion. With respect to factor analysis, IRT has two main advantages: It can cope with missing variables; therefore, the MLC method manipulations can be included as separate variables, allowing the separate analysis of each sample. Additionally, IRT explicitly allows for the separate variables to have different ranges. While some measures are better at distinguishing between risk-averse respondents, other measures are better at distinguishing between risk-seeking respondents. Therefore, IRT will give more precise estimates of the latent variable than a factor-weighted or an unweighted composite score, since, for each respondent, most of the weight is placed on the measures that are most relevant to the respondent's domain. Including the four manipulations of the MLC method, seven methods in total are included in the IRT analysis. The result of this analysis is an empirical Bayes mean predicted value for the latent variable theta (θ). Because all four elicitation methods have strong correlations to the latent variable, this variable can be assumed to be risk aversion, since all four methods are designed to measure risk aversion. For a full description of the graded response model and IRT, see Cohen and Kim (1998), Embretson and Reise (2013), and Van der Linden and Hambleton (2013).

Two outputs from IRT are used. First, the discrimination coefficient indicates the extent to which a question can distinguish between different levels of RRA. Second, the estimates of the difficulty coefficients provide the range in which the measure can reliably estimate RRA. Measures with a high discrimination coefficient can very accurately distinguish between risk aversion levels within the difficulty range. However, given the amount of options within the measure, a higher discrimination coefficient normally means a smaller difficulty range. Risk aversion levels that are higher than the highest difficulty or lower than the lowest difficulty cannot be reliably estimated. This result implies that a trade-off exists and that, given the amount of options (i.e. choices), measures can be either precise over a small range or less precise over a larger range.

2.6 Framing effects in the MLC method

In addition to creating a composite score, I assess the individual methods, particularly the framing effects in the MLC question. Therefore, the respondents were distributed in four samples, each being presented a specific manipulation of the MLC question (see Section 2.2). Given that the samples are selected randomly and that the sample size is sufficiently large, the differences between the four samples reflect differences in framing due to the manipulations. A Kolmogorov–Smirnov test for the equality of distributions is used to formally test whether these differences are significant and thus whether there are framing effects. Summary statistics show how manipulations influence the mean RRA results and their variance. Finally, differences in discriminative power and the range of covered values of theta (θ) resulting from IRT reflect how effective each manipulation is in eliciting normative risk preferences in the pension domain.

3 Data

3.1 Survey response

Data were collected with a dedicated online survey that was distributed to the participants in four pension funds, including the funds' retirees and the employees of five companies. A total of 34,477 participants (employees and retirees) were invited to take the survey; 9,891 clicked on the survey link, for a response rate of 28.7%.

Respondents were included in the analysis when they answered at least the MLC question, which required answering the preceding sociodemographic information questions. A total of 656 (6.6%) respondents did not answer this question and were eliminated from further analysis. Another 1,315 (13.3%) respondents had at least one missing answer but are included in the analysis as much as possible. Comparing this group with the group that completed the entire survey shows that the incomplete survey group is slightly less educated and younger and has a lower income. The results are thus not fully representative of the population that accepted the invitation or, likely, the population as a whole. However, this is not problematic, since heterogeneity in responses, rather than representativeness, is important for this research. Table 2 shows the number of participants and respondents and the response rates for the different pension funds and for active members and retirees. The response rates range from 23% to 52% for the different pension funds. Generally, the response rates are a few percentage points higher for the active population than for the retired population.

Table 2: Response rates

Plan	Population			Response			Response rate		
	Active	Retired	Total	Active	Retired	Total	Active	Retired	Total
1	18,058	7,366	25,424	4,972	1,654	6,626	27.5%	22.5%	26.1%
2	2,999	1,568	4,567	966	371	1,337	32.2%	23.7%	29.3%
3	3,137	477	3,614	658	165	823	21.0%	34.6%	22.8%
4	718	154	872	379	70	449	52.8%	45.5%	51.5%
Total	24,912	9,565	34,477	6,975	2,260	9,235	28.0%	23.6%	26.8%

Summary data are presented in Table 3. The pension funds participants differ extensively in terms of average age, gender, average income, and average level of education. Plan 1 has the largest active population and the lowest average income and plan 4 has the youngest active population, the highest level of education, and the highest average income. There are also notable differences between the active population and the retired population. Generally, retired participants are more often male and have a higher income and a slightly lower level of education.

Table 3: Summary data

Plan	Avg. Age		% Male		Avg. Income		Avg. Education	
	Active	Retired	Active	Retired	Active	Retired	Active	Retired
1	50.1	69.9	56.5%	80.1%	2.636	3.057	4.56	4.05
2	44.9	69.8	78.5%	90.0%	2.975	3.930	5.12	4.90
3	48.0	65.3	69.4%	71.8%	2.806	2.936	3.74	3.58
4	41.9	72.1	73.7%	81.4%	4.229	5.271	5.20	4.63
Total	48.8	69.6	61.7%	81.1%	2.785	3.260	4.59	4.18

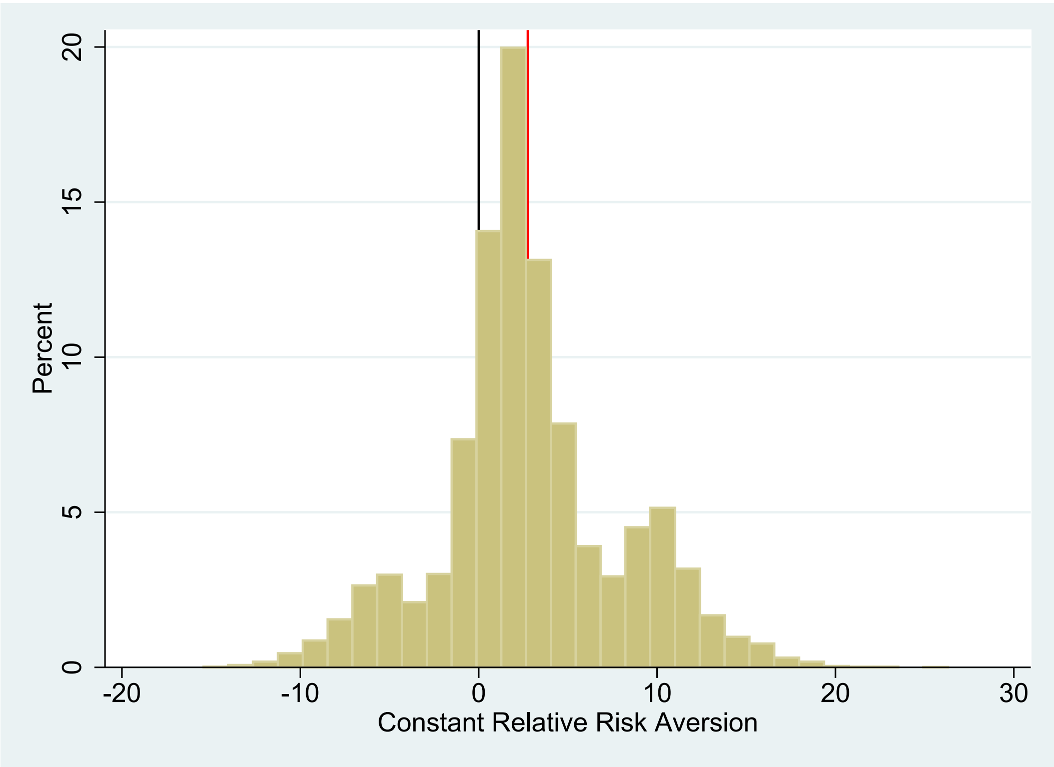
3.2 Elicitation methods

The results for the MLC method are presented in Table 4. This table shows the point at which the respondents switched from one pension plan (lottery) to the other, dependent on the specific manipulation the respondent faced (Table 1). Manipulations 2 to 4 were assigned to 10% of the participants each, while the first manipulation was presented to the remaining 70%. Due to non-response and dropouts, the response rates differ slightly across manipulations. Because there are no substantial differences in dropouts for or just after the MLC question, these should not influence the results. Using the method described in Section 2.2, the responses are transformed into a distribution of RRA. The resulting distribution, depending on the manipulation, is presented in Figure 2; RRA is distributed between -15 and 26, with a mean of 2.84.

Table 4: Responses for the MLC sample

Switching point (p =)	Sample 1 *Standard*	Sample 2 *Alternate values*	Sample 3 *Opposite direction*	Sample 4 *Different starting value*
10%	648 (9.9%)	174 (19.9%)	55 (6.2%)	-
30%	219 (3.3%)	51 (5.8%)	41 (4.6%)	-
50%	686 (10.4%)	155 (17.8%)	100 (11.3%)	172 (18.9%)
65%	696 (10.6%)	115 (13.2%)	241 (27.2%)	94 (10.3%)
80%	1,740 (26.5%)	173 (19.8%)	152 (17.2%)	251 (27.5%)
90%	887 (13.5%)	83 (9.5%)	96 (10.8%)	127 (13.9%)
95%	295 (4.5%)	19 (2.2%)	32 (3.6%)	52 (5.7%)
99%	415 (6.3%)	29 (3.3%)	64 (7.2%)	61 (6.7%)
100%	979 (14.9%)	74 (8.5%)	104 (11.8%)	155 (17.0%)
Total	**6,565**	**873**	**885**	**912**

Figure 2: Distribution of RRA



Notes: Distribution of RRA following from the MLC method. The black line represents risk neutrality ($RRA = 0$) and the red line represents the mean RRA.

The responses to the two seven-point Likert scale self-description questions are presented in Table 5. The responses to the simplified portfolio choice are presented in Table 6. All four methods have strongly heterogeneous results.

Table 5: **Results of the self-description method**

Result	**Stated_aversion**		**Careful**	
	Freq.	Percent	Freq.	Percent
1 (most risk seeking)	209	2.61	129	1.62
2	335	4.19	481	6.03
3	1,189	14.86	880	11.03
4	1,349	16.86	1,763	22.09
5	1,426	17.82	1,740	21.80
6	2,149	26.86	2,115	26.50
7 (most risk averse)	1,343	16.79	873	10.94
Total	8,000	100.00	7,981	100.00

Table 6: **Results of the portfolio choice method**

% allocation to bonds	**Freq**	**Percent**
0	130	1.54
10	47	0.56
20	204	2.42
30	334	3.96
40	409	4.85
50	1,214	14.39
60	739	8.76
70	1,435	17.01
80	1,350	16.00
90	1,377	16.32
100	1,197	14.19
Total	8,436	100.00

4 Measuring normative risk preferences

As previous research has frequently shown, elicited risk preferences tend to depend on the measure of elicitation used (e.g. Harrison et al. (2007); Kapteyn and Teppa (2011)). Therefore, to measure normative risk preferences as closely as possible, one should combine multiple relevant measures of risk preferences (Kapteyn and Teppa, 2011; Menkhoff and Sakha, 2016). By analysing the joint correlation of different methods, one eliminates measurement noise and method-specific biases and the normative value of elicited risk preferences increases. Principle component factor analysis can be used to study the joint correlation of risk preferences measured with different elicitation methods.

4.1 Principle component factor analysis

The results of the principle component factor analysis are presented in Table 7. Two factors are retained in this factor analysis, with eigenvalues of 2.18 and 1.21 respectively. For the first factor, four variables, which are the four different risk preference variables, have a loading higher than 0.4. Since the primary factor variable scores for all four methods designed to measure risk aversion, this factor is now identified as the latent variable risk aversion. The second factor has two variables with a loading higher than 0.4: education and self-estimated financial literacy. This factor is identified as financial/questionnaire literacy.

Table 7: Principal component analysis

Variable	Factor 1	Factor 2	Uniqueness
MLC (RRA)	0.6799	0.2632	0.4685
Stated aversion	0.8482	-0.1964	0.2420
Allocation to bonds	0.8019	-0.2648	0.2868
Careful	0.4363	0.1913	0.7730
Education	-0.1545	0.7095	0.4728
Financial literacy	-0.1509	0.7172	0.4629

Notes: This table shows the results of the CPA analysis with varimax rotation and two retained factors (eigenvalues greater or equal to one). The first factor is identified as risk aversion and the second factor as financial literacy. Variable loadings greater or equal to 0.4 are in gray.

All four risk aversion variables have a loading higher than the threshold, indicating that all four variables add relevant information about the latent variable risk aversion. In addition, the risk aversion variables do not have loadings greater or equal to 0.4 for the second factor, which shows that financial literacy has only a limited effect on these risk preference elicitation methods. However, the loadings are substantially different from zero, with both positive and negative signs. So, financial literacy does influence individual risk preference elicitation methods, but in different directions, stressing the importance of combining multiple elicitation methods.

4.2 IRT

The loadings on the two factors of the principle component analysis show that the four risk preference elicitation methods applied, adequately measure the same latent variable, identified as risk aversion. Although individual measures are influenced by the cognitive abilities of the respondents, the results show that combining the available measures gives a more reliable estimate of risk aversion, because measurement noise and individual elicitation method biases - which go both ways for different methods - are reduced.

Therefore, to estimate the composite score (of the latent variable θ), the results of our four measures are combined by applying IRT. Then, IRT can be used to estimate the latent variable (θ) using multiple measures with different locations (the range covered of the latent variable) and discrimination (measure precision).

Table 8: IRT results

Measure & Manipulation	Discrimi-nation	Difficulty (range) min	max	Correlation composite score all measures	minus measure#
MLC	1.004	-2.596	3.009	0.484	0.400
Standard	(0.031)	(0.078)	(0.089)		
MLC	0.936	-1.786	3.600	0.470	0.382
alternate values	(0.086)	(0.161)	(0.329)		
MLC	0.892	-3.746	2.534	0.455	0.357
Opposite direction	(0.076)	(0.321)	(0.221)		
MLC	1.061	-1.681	2.749	0.544	0.416
Different starting value	(0.084)	(0.133)	(0.208)		
Stated aversion	6.061	-1.999	0.966	0.969	0.710
	(0.767)	(0.039)	(0.022)		
Careful	0.479	-8.835	4.494	0.239	0.205
	(0.024)	(0.457)	(0.224)		
Allocation to bonds	2.269	-2.763	1.346	0.791	0.662
	(0.069)	(0.060)	(0.027)		

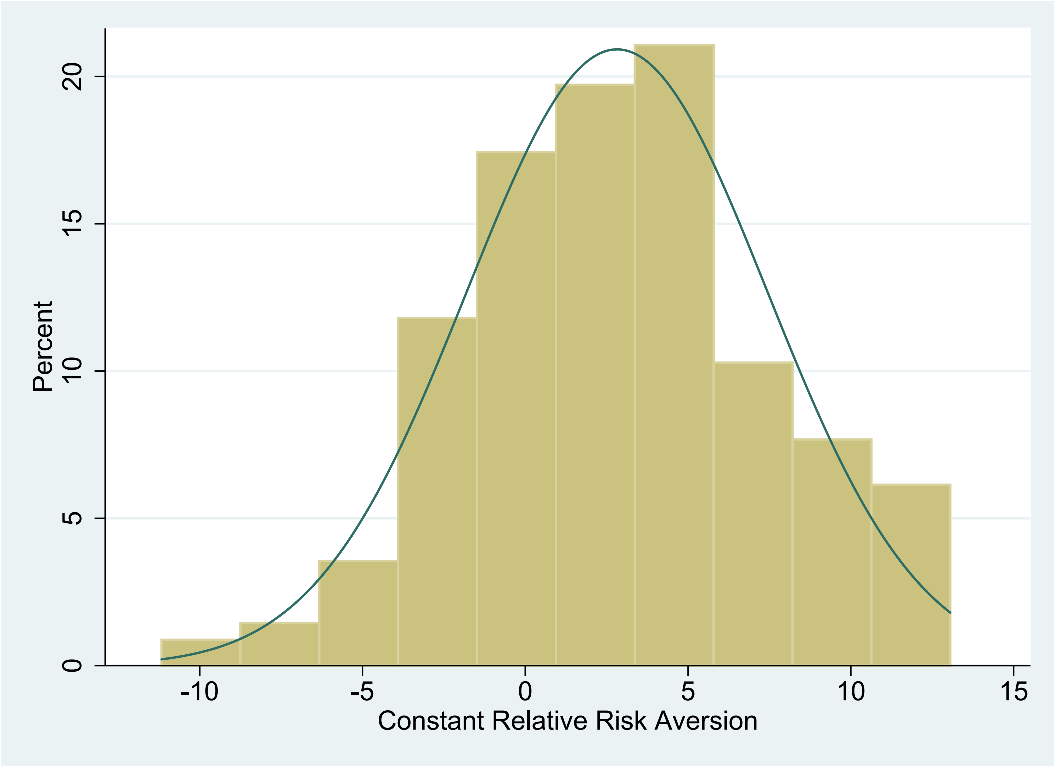
Notes: This table shows the results for IRT. The number of observations is 9,235. # indicates correlation with a composite score composed of all the other measures.

The results of the IRT analysis in Table 8 show relatively minor changes in the location and discriminative values of the different manipulations of the MLC method. The stated aversion and allocation to bonds measures have greater precision but (therefore) also cover a smaller range of risk aversion. The careful measure has the least precision and the largest range of risk aversion covered.

The results of our IRT analysis can be used to estimate the latent variable. These estimations (θ) are rescaled to the domain of RRA using the distribution of the MLC method (linear transformation). However, due the reduction of noise following from the combination of methods, the standard deviation of RRA is reduced by 9.89%. For the composite score, RRA is therefore distributed with a mean of 2.82 and a standard deviation of 4.62. The estimates are presented in the histogram in Figure 3.

The estimate θ has a high correlation with all the measures except for the self-description variable *Careful*. The correlation coefficients decrease only slightly if the correlation is analysed for the composite score composed of all measures except for the measure under analysis, indicating the robustness of the results. The correlation coefficients show that both adjustment to the relevant domain and keeping the questions as simple as possible increase the normative value of the results (a higher correlation with θ). However, in application to financial problems, some quantitative value of risk aversion is required, necessitating more quantitatively oriented (and therefore more difficult) methods, such as the MLC question. The best manipulation of this specific method is analysed in Section 5.

Figure 3: Estimates of the latent variable θ



Notes: This figure shows the histogram of the estimates for the latent variable for risk aversion θ. The values are rescaled to the domain of RRA using the results of the MLC method.

4.3 Predicting normative risk preferences

Regressing the augmented risk preferences on sociodemographic information and pension fund membership provides insight into the extent to which individual pension fund participants' risk preferences can be estimated. This is relevant information, since it could avoid the need for eliciting risk preferences. If risk preferences can be predicted closely enough, delegated decision makers can save on costly risk preference elicitation. In addition, the normative value of the results can be assessed by analysing the regression results of risk aversion following the MLC method and risk aversion following IRT (the composite score). If the signs and sizes of the coefficients are in line with earlier findings, they are more reliable and the normative value of the results are higher. The results of the regression analyses are presented in Table 9.

First, regarding the normative value of the risk aversion results, the coefficients from model 2 (composite score) are mostly larger and more often significant than for model 1 (MLC). The coefficients of income and home ownership (higher income/owns a house, less risk averse) and education (higher education, less risk averse) are in line with earlier findings (Harrison et al., 2007; Holt and Laury, 2002) and significant at the 5% level for the composite score but not for MLC risk aversion. In addition, the predictive value (R^2) is far higher for the composite (13.5%) score than for the MLC method (1.7%). These findings suggest that the risk aversion from the composite score is more reliable and thus has a higher normative value than MLC risk aversion.

Although the predictive power of the composite score (13.5%) is far higher than for the MLC method, it is not high enough to replace the elicitation of individual risk preferences. The vast majority of variation remains unexplained, for example, because of genetic variation (Zyphur et al., 2009), and elicitation is necessary for a reliable estimate of risk aversion. On average, the results indicate that younger persons, persons with a higher income or who own a house, men, and higher-educated persons are less risk averse, which is in line with earlier findings (Harrison et al., 2007; Holt and Laury, 2002; Jianakoplos and Bernasek, 1998; Yao et al., 2011). Although older individuals are more risk averse, retirement has no significant effect. Being retired (less human capital) therefore does not seem to influence risk aversion.

Table 9: Regression analysis results

Variables	(1) **Risk aversion** MLC		(2) **Risk aversion** Composite score	
Age	0.031***	(0.006)	0.042***	(0.005)
Retired	-0.185	(0.183)	-0.081	(0.152)
Income ($1,000)	-0.068	(0.054)	-0.361***	(0.049)
House	0.129	(0.166)	-0.333**	(0.130)
Gender:[a]				
Female	-0.972	(0.594)	-0.576	(0.460)
Male	-1.867***[b]	(0.592)	-2.565***[b]	(0.459)
Education:				
Pre-vocational education	-1.553	(1.256)	-1.541	(0.987)
Secondary vocational education	-0.837	(1.251)	-1.729*	(0.984)
Senior general secondary education	-0.857	(1.248)	-2.054**	(0.984)
Professional education	-0.939	(1.244)	-2.310**	(0.981)
Academic education	-0.865	(1.247)	-3.086***	(0.985)
Pension fund:				
2	0.567***	(0.143)	0.051	(0.133)
3	0.752***	(0.202)	0.401**	(0.163)
4	-1.258***	(0.233)	-3.246***	(0.235)
Constant	3.732***	(1.409)	6.177***	(1.114)
R^2	0.017		0.135	

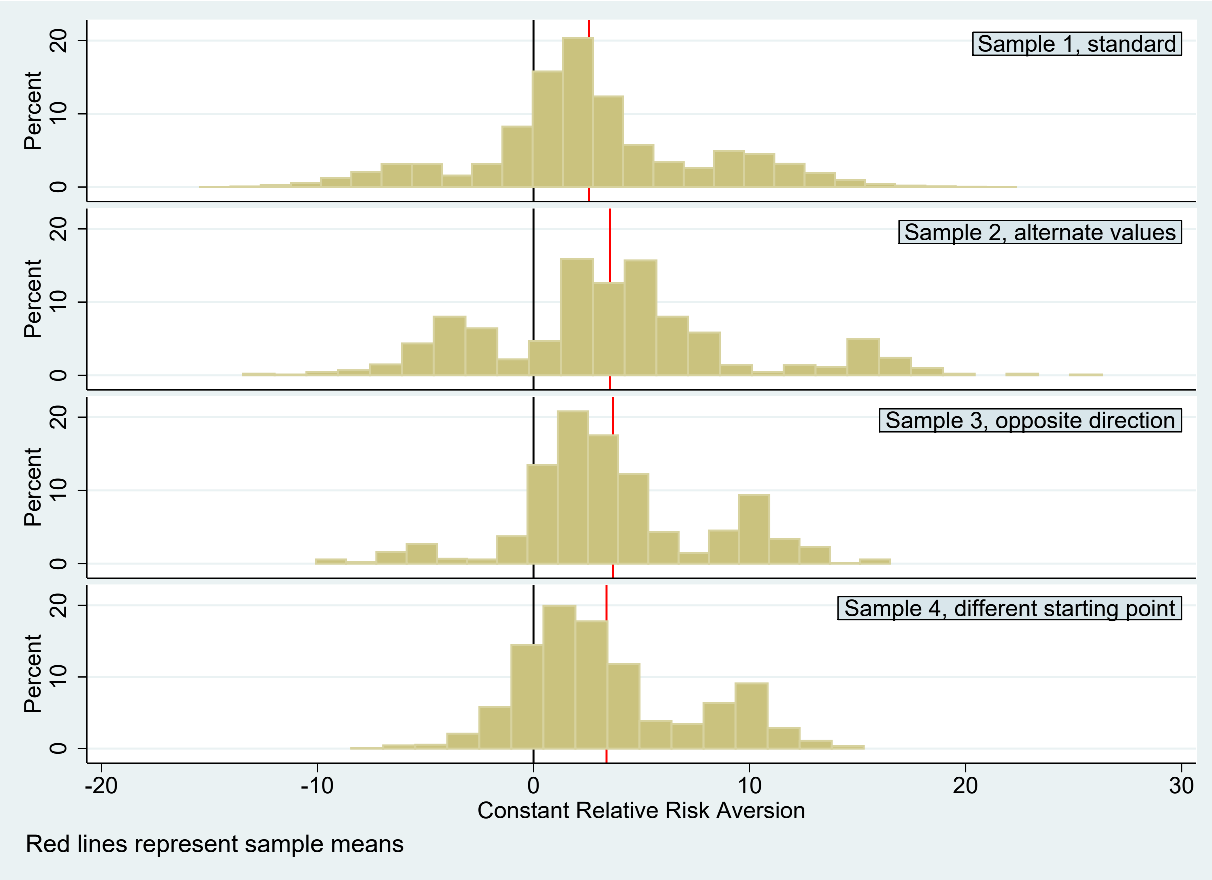
Notes: This table shows the results of the regression analysis of observable characteristics on RRA, with robust standard errors. The superscript [a] indicates compared to the group ($N = 98$) that did not report gender; [b] indicates significantly different ($p < 0.01$) from gender being female. Standard errors are in parentheses, N = 9,235, * p<0.10, ** p<0.05, *** p<0.01.

5 Framing effects in the MLC method

Now that the latent variable for risk aversion has been estimated in the previous section, the results can be used to assess the reliability of the different manipulations of the MLC method. The respondents are randomly assigned to one of four samples, each of which received the MLC question in a distinct framing. The distribution of the implied RRA resulting from the MLC question, given framing, is presented in Figure 4.

Figure 4 clearly shows that changes in the framing of the question influence the implied distribution of RRA. The mean value of RRA (indicated by the red line in Figure 4) is 2.56 for the baseline question, 3.54 for the question with alternate values, 3.68 for the reversed question, and 3.38 for the question with a later starting point. Also, the standard deviation changes with manipulation (5.19, 6.00, 4.41 and 4.03 respectively).

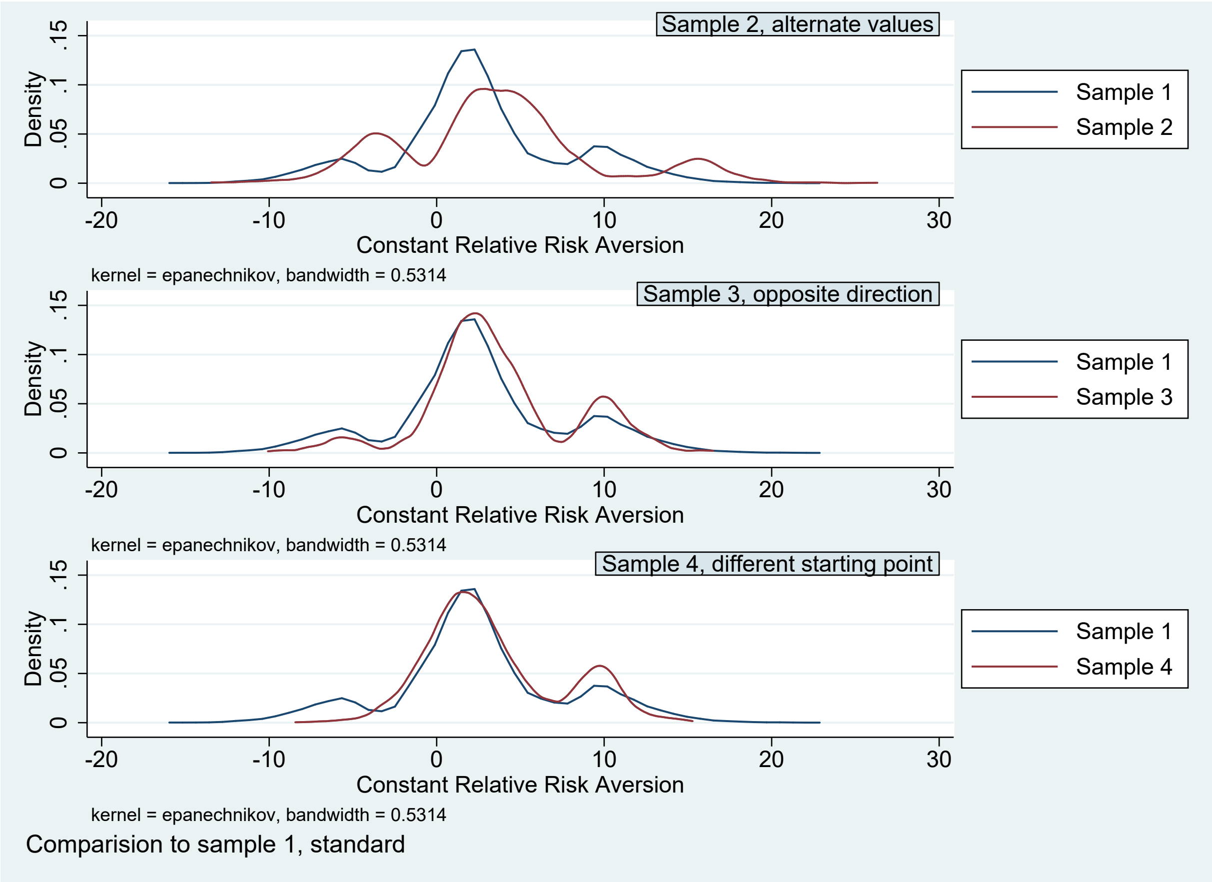
Figure 4: Distribution of RRA, dependent on manipulation



Notes: This figure shows the histograms of constant RRA for each sample. The black line represents risk neutrality ($RRA = 0$) and the red line represents the sample mean.

A Kolmogorov–Smirnov test for the equality of distribution functions shows that the latter three samples differ significantly ($p = 0.000$) from the first sample. Figure 5 presents the kernel density estimates for the latter three samples compared to the first sample. The largest differences are between the sample with alternate amounts (sample 2) and the baseline sample (sample 1). This manipulation results (on average) in higher values of risk aversion, which are also more widely distributed (greater variation). The effects of the other two samples are smaller, despite the results shifting towards higher risk aversion, the overall form of the distribution is quite similar across samples.

Figure 5: Kernel density estimates



Notes: Kernel density estimates for constant RRA for each sample

As the previous analysis showed, applying manipulations to the MLC method results in significantly different RRA results. This method is thus not impervious to manipulations in amounts, probabilities, and sequence. However, so far, it remains unclear what manipulation of the MLC method provides the most reliable elicitation method or, put differently, which manipulation yields values closest to (the estimation of) the latent variable normative risk aversion.

Analysing the correlation between the results of the MLC method and the composite score shows the extent to which the MLC method corresponds to the estimation of the latent variable risk aversion. The higher the correlation, the more valuable the method is in estimating true risk preferences. Table 10 shows the correlation with the composite score and properties resulting from the IRT analysis. Sample 4, a different starting point, has the highest correlation and the highest discrimination coefficient. This measure thus allows for a relatively reliable estimation of risk aversion. However, due to the high discrimination and limited number of options, the range of this manipulation is limited. The standard question (sample 1) has both a relatively high correlation and discrimination and, compared to the fourth sample, a larger range. Samples 2 and 3, despite having large ranges, are less reliable and therefore not optimal for measuring risk preferences. The best measure (sample 1 or 4) depends on the characteristics of the preferences measured. Since pension domain risk preferences are normally more risk averse (Van Rooij et al., 2007), the different starting point manipulation (sample 4) is selected as the preferred measure, since its range better covers pension domain risk preferences.

Table 10: Framing effects

Manipulation		**Correlation**	**IRT**	
			Coefficient	**Range**
1	Standard	0,484	1.004	-2.596 - 3.009
2	Alternate values	0,470	0,936	-1.786 - 3.600
3	Opposite direction	0,455	0,892	-3.746 - 2.534
4	Different starting point	0,544	1,061	-1.681 - 2.749

Notes: This table shows the properties of different manipulations of the MLC questions following from the correlations and IRT analysis.

6 Base pension framing

Another type of framing that was included in the survey was the inclusion or exclusion of the base pension. The base pension in the Netherlands (2016) is between \$862 and \$1,696 a month, depending on the presence and situation of a partner. For low- to medium-income groups, this is a substantial part of the total retirement income (40–79% of the expected retirement income for median-income groups) and, since this base pension is almost risk free, this should have consequences for the risk taken in one's occupational (second pillar) pension. More specifically, for the distribution of total retirement income to be in line with risk preferences, occupational pension income should be invested more aggressively in the presence of a base pension, to offset the lack of risk in this part of retirement income.

For two of the four pension funds, half the population is presented the entire survey in the context of only occupational pension (excluding the base pension). This creates two groups that differ only in the extent of the framing of the questions with or without the base pension. Table 11 shows the results of the estimation of pension income and the evaluation of the participants' different pension incomes. On average, the participants expect to receive an extra \$434 (monthly, net of taxes) with the inclusion of the base pension. Comparing this to a (gross) base pension between \$862 and \$1,696 shows that the participants, on average, do take base pension into account, but imperfectly. In addition, the base pension is not considered to be risk free by everyone, since the differences between both samples increase with more favourable estimations. Participants who have a partner exhibit, on average, smaller differences, in line with lower base pensions for married individuals.

Table 11: Pension income expectations

Estimated pension:	**Base pension**			**Partner**	**No partner**
	With	Without	Diff.	Diff.	Diff.
Expected	2,259	1,825	434	366	480
Very low	1,495	1,157	338	331	358
Low	1,723	1,374	349	334	390
Neither low nor high	1,993	1,633	360	340	417
High	2,260	1,887	373	345	449
Very high	2,638	2,260	378	313	590

Notes: This table shows the expectations of pension income, with and without a base pension, the difference between both forms of framing, and the difference for those with and without a partner.

The effect of framing questions with or without the base pension on elicited risk preferences can be estimated with regression analysis. The results of the MLC question are explained using regression analysis for the full model of explanatory variables (Table 9) and a dummy is included to indicate the exclusion of a base pension (i.e. inclusion acts as the reference). This analysis shows that excluding the base pension increases RRA (from the MLC question) by 0.269 ($p = 0.028$).[1] Although the effect size is relatively weak, it is counterintuitive. Since individuals have a base pension (outside of the scope of the study), this would suggest that individuals can take more risk with their pension income; however, the opposite effect is found. Both in the case of the counterintuitive effect or in the case of no effect, the shifts in observed risk preferences are not in line with normative risk preferences. Given this finding, it is important that the survey reflect the perspective of the respondents as closely as possible, because they have difficulty processing different situations in their revealed preferences. In the case of pension domain risk preferences, this means eliciting risk preferences for total retirement income, net of taxes, since this is closest to the participants' perception.

7 Conclusion

I use a combination of risk preference elicitation methods and manipulations to analyse risk preferences in the pension domain. Since different methods tend to give different results, I combine the results of different methods into a composite score using IRT. Principle component factor analysis shows that the four methods applied describe, to various degrees, a common latent variable, which is identified as risk aversion. Comparing the composite score to the separate methods shows that the composite score is less affected by measurement noise and method-specific biases. In addition, the explanatory power of several well-studied sociodemographic characteristics (e.g. income, age, and gender), is greater for the composite score.

Using the composite score as a proxy for normative risk preferences, I assess the effect of four manipulations of the MLC method. All four methods influence the resulting distribution of risk aversion, but the

[1]Including interaction with income did not produce significant results

effect of changing the direction or the starting point of the questions has a smaller effect (shift towards risk aversion) than changing the amounts (shift towards risk aversion and larger variation). Given that individuals tend to be more risk averse in the pension domain, a later starting position increases the normative value of the results.

A second manipulation, the exclusion of a base pension, shows that respondents incompletely take the relevant factor into account and risk preferences are not adjusted appropriately. Therefore, risk preferences should be elicited as closely as possible to the practical situation, thus including a base pension in the case of retirement income.

The results show that revealed risk preferences differ with different elicitation methods or with manipulations of elicitation methods. The revealed risk preferences are therefore not equal to normative risk preferences, that is, the risk preferences that represent an individual's true interests. Using a combination of augmented measures, including framing the methods closely to the relevant real-life situation, improves the reliability of the results and therefore provides the best estimation of normative risk preferences.

References

Alserda, G. A. G., B. G. C. Dellaert, L. Swinkels, and S. G. Van der Lecq (2017). The measurement of risk preferences heterogeneity and the impact on optimal asset allocation. *mimeo, Erasmus School of Economics, Rotterdam*.

Anderson, L. R. and J. M. Mellor (2009). Are risk preferences stable? comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty 39*(2), 137–160.

Becker, G. M., M. H. De Groot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral science 9*(3), 226–232.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2008). How are preferences revealed? *Journal of Public Economics 92*(8), 1787–1794.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2013). Simplification and saving. *Journal of economic behavior & organization 95*, 130–145.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics 67*(1), 41–80.

Cohen, A. S. and S.-H. Kim (1998). An investigation of linking methods under the graded response model. *Applied Psychological Measurement 22*(2), 116–130.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature*, 448–474.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty 41*(3), 219–243.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

EIOPA (2013). Survey of eu practice on default investment options. Technical Report 13/033, Frankfurt, European Insurance and Occuational Pensions Authority.

Embretson, S. E. and S. P. Reise (2013). *Item response theory*. Psychology Press.

Fellner, G. and B. Maciejovsky (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology 28*(3), 338–350.

Frijns, J. (2010). Dutch pension funds: aging giants suffering from weak and inconsistent risk management. *Rotman International Journal of Pension Management 3*(2), 16–23.

Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 631–645.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal 5*(2), 41–53.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2007). Estimating risk attitudes in Denmark: A field experiment. *The Scandinavian Journal of Economics 109*(2), 341–368.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Hsee, C. K. and E. U. Weber (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making 12*.

Iyengar, S. S., G. Huberman, and W. Jiang (2004). How much choice is too much? contributions to 401 (k) retirement plans. *Pension design and structure: New lessons from behavioral finance*, 83–95.

Iyengar, S. S. and E. Kamenica (2006). Choice overload and simplicity seeking. *University of Chicago Graduate School of Business Working Paper 87*.

Jianakoplos, N. A. and A. Bernasek (1998). Are women more risk averse? *Economic Inquiry 36*(4), 620–630.

Kapteyn, A. and F. Teppa (2011). Subjective measures of risk aversion, fixed costs, and portfolio choice. *Journal of Economic Psychology 32*(4), 564–580.

LeBoeuf, R. A. and E. Shafir (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making 16*(2), 77–92.

Levin, I. P. and S. S. Hart (2003). Risk preferences in young children: Early evidence of individual differences in reaction to potential gains and losses. *Journal of Behavioral Decision Making 16*(5), 397–413.

Lévy-Garboua, L., H. Maafi, D. Masclet, and A. Terracol (2012). Risk aversion and framing effects. *Experimental Economics 15*(1), 128–144.

Lusardi, A. and O. Mitchell (2007). Financial literacy and retirement preparedness: Evidence and implications for financial education. *Business economics 42*(1), 35–44.

March, J. G. and Z. Shapira (1992). Variable risk preferences and the focus of attention. *Psychological review 99*(1), 172.

Menkhoff, L. and S. Sakha (2016). Estimating risky behavior with multiple-item risk measures: An empirical examination.

Rozinka, E. and W. Tapia (2007). Survey of investment choice by pension fund members. Technical report, OECD Publishing.

Soane, E. and N. Chmiel (2005). Are risk preferences consistent?: The influence of decision domain and personality. *Personality and Individual Differences 38*(8), 1781–1791.

Van der Linden, W. J. and R. K. Hambleton (2013). *Handbook of modern item response theory*. Springer Science & Business Media.

Van Rooij, M. C., A. Lusardi, and R. J. Alessie (2011). Financial literacy and retirement planning in the Netherlands. *Journal of Economic Psychology 32*(4), 593–608.

Van Rooij, M. C. J., C. J. M. Kool, and H. M. Prast (2007). Risk-return preferences in the pension domain: are people able to choose? *Journal of Public Economics 91*(3), 701–722.

Viceira, L. M. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *The Journal of Finance 56*(2), 433–470.

Weber, E. U., A.-R. Blais, and N. E. Betz (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making 15*(4), 263–290.

Yao, R., D. L. Sharpe, and F. Wang (2011). Decomposing the age effect on risk tolerance. *The Journal of Socio-Economics 40*(6), 879–887.

Zyphur, M. J., J. Narayanan, R. D. Arvey, and G. J. Alexander (2009). The genetics of economic risk preferences. *Journal of Behavioral Decision Making 22*(4), 367–377.