Matteo M. Galizzi, Sara R. Machado and Raffaele Miniaci

Temporal stability, cross-validity, and external validity of risk preferences measures: experimental evidence from a UK representative sample

Working paper

Temporal Stability, Cross-Validity, and External Validity of Risk Preferences Measures: Experimental Evidence from a UK Representative Sample

Matteo M. Galizzi*[1], Sara R. Machado[2] and Raffaele Miniaci[3]

[1] London School of Economics: Department of Social Policy, LSE Health, LSE Behavioural Research Lab; OLD 2.35, Old Building, Houghton Street, WC2A 2AE London (UK). Email: m.m.galizzi@lse.ac.uk
[2]London School of Economics: Department of Social Policy, LSE Health; Old Building, Houghton Street, WC2A 2AE London (UK). Email: s.machado@lse.ac.uk
[3]University of Brescia: Department of Economics and Management; Via San Faustino 74b, 25100 Brescia (Italy). Email: raffaele.miniaci@unibs.it

12 August, 2016

Abstract

We conduct an "artefactual" field experiment to incorporate three different risk preferences measures within the Innovation Panel (IP) of the UK Household Longitudinal Survey (UKHLS). We randomly allocate to an experimental module a nationally representative sample of 661 adult respondents to the IP Wave 6 (IP6). These subjects respond to the incentive-compatible tasks by Holt and Laury (2002) (HL), and by Binswanger (1980, 1981) and Eckel and Grossman (2008) (B-EG), and to the SOEP survey questions by Dohmen et al. (2011) for self-reported willingness to take risks in general (SOEP-G), in finance (SOEP-F), and in health (SOEP-H). One year later (IP7) the same measures are repeated for 413 of these respondents. This design allows us to systematically test, for a UK representative sample, the validity of the three measures along three dimensions. First, we look at cross-validity by testing how responses at one point in time correlate across the three tasks, assuming a Constant Relative Risk Aversion (CRRA) utility function. Second, we look at temporal stability by comparing the responses across IP6 and IP7. Third, we look at external validity by considering a range of risky health and financial behaviors in the UKHLS. We have three main findings. First, concerning cross-validity, we find evidence that the different measures generally correlate and map into each other, although their associations are not perfect. Second, concerning temporal stability, there are significant and positive correlations of the B-EG, HL, and SOEP measures across IP6 and IP7. Finally, we find mixed evidence concerning external validity.

Keywords: field experiments; risk aversion; behavioural data linking; health behaviors.

JEL Codes: C83, C93, D81, I12

*Corresponding author.

1 Introduction

This study takes a systematic look at the stability and validity of different risk preferences measures using a longitudinal sample of adults, representative of the UK population.

The stability of risk preferences measures can be assessed at two levels: over time (*temporal stability*), and across different elicitation methods (*cross-validity*).[1] Besides cross-validity, the validity of risk preferences measures can be assessed in terms of *external validity*, that is, by looking at the extent to which they associate with, and are able to predict, a range of field behaviors. We investigate the issue of stability and validity of risk preferences measures along these three directions.

We conduct an artefactual field experiment (in the sense of Harrison and List, 2004) with a UK representative sample of adult respondents within the Innovation Panel (IP) of the UK Household Longitudinal Survey (UKHLS, also known as *Understanding Society*). In the sixth wave of the Innovation Panel (IP6) we randomly select a sub-sample of 661 respondents and we elicit their risk preferences using two experimental methods with real monetary rewards: the multiple price list (MPL) binary-lotteries method by Holt and Laury (Holt and Laury, 2002; HL), and the ordered lottery sequence method originally proposed by Binswanger (Binswanger, 1980, 1981), and then applied by Eckel and Grossman (Eckel and Grossman, 2008; B-EG). The same individuals are asked to answer the survey questions incorporated in the German Socio-Economic Panel survey (SOEP) and in other representative surveys across the world (Rieger et al., 2014; Becker et al., 2015; Josef et al., 2016). These survey questions record the respondents' self-reported risk attitudes in general (SOEP-G), in financial matters (SOEP-F), and in health matters (SOEP-H), using Likert scales. One year later, in the Innovation Panel wave 7 (IP7) we repeat the three risk preference measures for 413 of the original participants to the experiment.

The within-subjects design of the experiment allows us to investigate the *cross-validity* of the risk preferences measures by comparing, at an individual level, the responses across the experimental methods - and their implied coefficients of risk aversion - as well as with the responses to the survey questions.

We then exploit the longitudinal nature of the Innovation Panel to directly look at the *temporal stability* of the three different measures of risk preferences.

Finally, we take advantage of the wealth of information provided by the UKHLS data to systematically assess the *external validity* of the different measures of risk preferences with respect to a range of health- and finance-related risky behaviors.

To the best of our knowledge, our study is the first to look jointly and systematically at the three dimensions of cross-validity, temporal stability, and external validity of two experimental measures of risk preferences (B-EG and HL) and three risk attitudes questions (SOEP-G, SOEP-F, and SOEP-H) using a nationally representative sample. Looking systematically and comprehensively at all the three validity dimensions together is important

[1] Andersen et al. (2008b) distinguish between *conditional* stability and *unconditional* temporal stability: see Section 2.2 for a more detailed discussion. Another, conceptually distinct, question relates to the issue of stability of risk preferences across different life domains (*domain-specificity*): see, for instance, Hanoch et al. (2006); Barseghyan et al. (2011); Einav et al. (2012), and Galizzi et al. (2016c) for a recent summary of the literature looking at the domain-specificity of risk preferences.

because it allows us to draw conclusions on the *overall validity* of different risk preferences measures that would otherwise be missed when looking at each aspect in isolation.

When looking at *cross-validity*, our study is close in design and concept to Dave et al. (2010), Dohmen et al. (2011), Loomes and Pogrebna (2014), Crosetto and Filippin (2015), and Vieider et al. (2015b), among other studies, but it is the first to look at two experimental measures of risk preferences and three risk attitudes questions for a UK representative sample. Our study, moreover, also looks at the temporal stability of all these measures, like Andersen et al. (2008a) and Harrison and Lau (2014) do for the HL measure in a representative sample in Denmark, and Josef et al. (2016) do for the SOEP measures in a representative sample in Germany. Finally, our study also systematically compares, for the first time, three different measures of risk preferences in terms of their external validity using a representative panel of survey respondents. Section 2 reviews these studies, and the rest of the literature, in greater detail, and discusses how our study builds on the existing body of evidence on each of the three validity aspects.

We have three main findings. First, concerning *cross-validity*, the different measures of risk preferences correlate and map into each other, although their associations are far from being perfect. In particular, the two experimental measures are correlated to each other; the HL experimental measure with high monetary stakes is significantly associated with the survey question about self-reported risk attitude in general, but not with the self-reported risk attitudes in the financial and health domains; the HL experimental measure with lower monetary stakes is not correlated with any of the questions on self-reported risk attitudes; the B-EG experimental measure is instead positively associated to all the survey questions about risk attitudes.

Second, for what concerns *temporal stability*, we find that, after one year, less than one third of the subjects choose the same preferred option in the lottery choices or in the self-reported risk attitude scales. Nonetheless, there is a significant and positive correlation between the risk aversion parameters implicit in the lottery choices in IP6 and IP7, as well as between the self-reported risk attitudes in the two waves. Overall, risk attitudes elicited through survey questions are more temporally stable than the experimental measures of risk preferences. Moreover, the parameters of risk aversion implicit in the B-EG and the HL lottery choices and the self-reported attitudes in IP6 significantly predict the corresponding risk aversion parameters and self-reported attitudes in IP7.

Third, we find mixed evidence concerning the *external validity* of the three risk attitude measures. In particular, none of the three measures is associated with the respondents' smoking status, their consumption of junk food, their likelihood of regularly saving, or the time horizons of their savings. The two experimental measures significantly associate with the subjects' BMI, whereas none of the survey questions do. Only the HL and the B-EG measure predict the regular consumption of fruit and vegetables and whether the respondents have a private pension fund, respectively; on the other hand, only the risk attitudes survey questions associate with the DOSPERT survey questions for risk-taking in hypothetical situations in finance and in health, and with heavy alcohol drinking.

Section 2 contains a brief discussion of the background evidence and related literature, while Section 3 discusses in detail the experimental design, setting, and tasks. Section 4 describes the theoretical framework and outlines the empirical analysis, while Section

5 presents the results. Section 6 looks more closely at the "inconsistent" subjects, while Section 7 briefly concludes.

2 Contribution to the existing literature

2.1 Cross validity

The psychology and economics literatures have developed a multitude of tests to measure risk attitudes and risk preferences. In psychology, risk attitudes are measured using the Balloon Analogue Risk Task (BART, Lejuez et al., 2002) and the Domain-Specific Risk-Taking Scale (DOSPERT, Weber et al., 2002; Blais and Weber, 2006). In experimental and behavioral economics, besides the already mentioned B-EG and HL methods, several other incentive-compatible methods have been proposed to measure risk preferences (see Cox and Harrison, 2008; Harrison and Rutström, 2008; Charness et al., 2013 for some recent reviews of the literature). Some examples are the Becker-DeGroot-Marschack mechanism, the First Price Auction and the Second Price Auction methods (Becker et al., 1964; Cox et al., 1988; Harrison, 1986, 1990; Isaac and James, 2000); the Allocation Task (AT) (Loomes, 1991); the Trade-Off Methods (Wakker and Deneffe, 1996; Barsky et al., 1997; Anderson and Mellor, 2009); the Charness-Gneezy-Potters investment game (Gneezy and Potters, 1997; Charness and Gneezy, 2010); the Lottery Panel test (Sabater-Grande and Georgantzis, 2002; Georgantzís and Navarro-Martínez, 2010); and, more recently, the Bomb Risk Elicitation Task (Crosetto and Filippin, 2013). Besides the HL method, other multiple price list (MPL) methods have been used, among others, by Harless and Camerer (1994); Hey and Orme (1994); Loomes and Sugden (1998); Donkers et al. (2001); Harbaugh et al. (2002, 2010); Harrison et al. (2005c, 2007a); Jacobson and Petrie (2009); Tanaka et al. (2010); Dohmen et al. (2011); Von Gaudecker et al. (2011); Sutter et al. (2013); Vieider et al. (2015a).

Despite the multitude of methods to measure risk preferences in lab and field settings, a relatively limited number of studies have directly looked at how these different measures correlate and map into each other, and the vast majority of these studies have considered student samples.

Table 1 summarizes the main features and findings of the existing studies that look at the cross-validity of different measures of risk preferences at one point in time.[2] With the exception of Dohmen et al. (2011) and Josef et al. (2016), no study has considered representative samples of the population to investigate cross-validity. Moreover, both Dohmen et al. (2011) and Josef et al. (2016) focus on the association between an experimental MPL measure (which is different from the HL method), and the self-reported willingness to take risk in general (SOEP-G).

The state-of-the-art evidence on the cross-validity is somewhat mixed and not conclusive. Among the studies that consider university student samples, Becker et al. (2015) compare SOEP measures with responses to MPL tests with real and hypothetical rewards and find significant correlations. The rest of the cross-validity studies conducted with

[2] All tables are reported in Appendix A.

student samples, however, provide more mixed evidence. Deck et al. (2013), for example, compare the HL test with a different experimental task based on the "Deal or no deal" game show, as well as with a survey measure, and find inconsistent responses across the methods, with most correlation coefficients being not statistically significant. Loomes and Pogrebna (2014) compare responses to the AT, B-EG, and HL tasks and find no correlation between any of them. Further evidence of imperfectly mapping responses across different methods to measure risk preferences in student samples has also been provided by Isaac and James (2000), Anderson and Mellor (2009), Reynaud and Couture (2012), Dulleck et al. (2013), Vieider et al. (2015b), and Attanasi et al. (2016) using a within-subject design, and by Crosetto and Filippin (2013) using a between-subject design. Dave et al. (2010) consider a general population sample in Canada, and compare responses across the B-EG and HL methods using experimental tasks that are similar to the ones we use: they find significantly higher risk aversion in the HL than in the B-EG task.[3]

As mentioned, only two cross-validity studies to date have considered representative samples of the population, both in Germany: Dohmen et al. (2011) and Josef et al. (2016) compare responses to an incentive-compatible MPL test with the SOEP-G measure for risk attitudes in general, and find small but significant associations.

We contribute to these existing studies by systematically assessing the cross-validity between two different state-of-the-art experimental measures of risk preferences (B-EG and HL) and three risk attitudes questions (SOEP-G, SOEP-F and SOEP-H), considering a representative sample of the UK adult population.

2.2 Temporal stability

Most of the existing evidence on the temporal stability[4] of experimental measures of risk preferences comes from studies with small, unrepresentative samples.[5] Evidence from larger samples and/or over longer periods of time is available for survey measures of risk attitudes or hypothetical gambles (e.g. Sahm, 2012; Guiso et al., 2013; Hoffmann et al., 2013; Josef et al., 2016). Recent reviews of the main studies assessing the temporal stability

[3]Evidence of imperfect mapping of risk preferences measures has also been provided by Charness and Viceisza (2015), considering a convenience sample in rural Senegal.

[4]Concerning temporal stability, Andersen et al. (2008b) distinguish between unconditional and conditional stability of risk preferences. *Unconditional stability* postulates that risk aversion literally remains constant over time. According to *conditional stability*, what remains constant over time is the function that links the risk aversion with the observable states of nature. Conditional stability is thus a weaker concept of stability of risk preferences. Conditionally temporally stable preferences are component of a broader approach often referred to as "state-dependent preferences" (Karni, 1983, 1987, 1993; Karni et al., 1983; Viscusi and Evans, 1990; Evans and Viscusi, 1991; Chambers and Quiggin, 2000).

[5] For instance, 30 students at Harvard University over a period of 10 weeks in Mosteller and Nogee (1951); 23 US farmers over a period of 2 years in Love and Robison (1984); 84 US businessmen over a period of 1 year in Wehrung et al. (1984); 80 students at University of York over a period of a week to 10 days in Hey and Orme (1994); 253 farmers in the Netherlands over a period of 1 year in Smidts (1997); 31 students at University of South Carolina over a period of 6 months in (Harrison et al., 2005b); 69 students at UCL over a period of 3 months in Vlaev et al. (2009); 141 MBA students at IESE Business School over a period of 3 months in (Baucells and Villasís, 2010); 73 undergraduate students at University of Munster over a period of 1 month in Zeisberger et al. (2012); 53 graduate students at University of Maastricht over a period of 5-10 weeks in Wölbert and Riedl (2013).

of risk preferences can be found in Zeisberger et al. (2012), and Chuang and Schechter (2015).

Three prominent studies in temporal stability of risk preferences in representative samples are the longitudinal studies by Andersen et al. (2008b), Harrison and Lau (2014), and Josef et al. (2016). Andersen et al. (2008b) measure risk preferences for a representative sample of 253 Danes using the HL method, and revisit 97 of them in four occasions, at a distance of up to 17 months after, and find that risk preferences are state contingent with respect to personal finances but are not changing over time. Josef et al. (2016) look at test-retest correlations for the SOEP survey questions in a large representative survey in Germany (around 11,000 subjects) over a period of up to ten years, and find moderate to high correlations (in a range of 0.42-0.58). The longitudinal artefactual field experiment by Harrison and Lau (2014) measure risk preferences for a representative sample of 413 Danes using the HL method and, one year after, repeat the same HL measures for 182 of those same subjects: once their structural estimations make full corrections for sample selection and attrition, they find evidence of temporal stability of risk preferences

With respect to this existing literature, our study considers a representative sample of the UK adult population to look systematically at the temporal stability of a battery of risk preferences measures, including two different experimental measures and three risk attitudes questions.

2.3 External validity

Concerning external validity, a large number of previous studies have looked at the associations between experimental measures for risk preferences and individual behaviors in a variety of contexts, from health to finance, from job choices to environmental behavior. A systematic review of this literature is beyond the scope of this study, but also on this third dimension the evidence to date is mixed and not conclusive.

Consider, for instance, a risky health behavior which is usually assumed to be associated with risk preferences: smoking behavior. A recent review of the literature on this topic can be found in Harrison et al. (2015a) and reveals that some studies find a significant association between smoking status and risk preferences (Reynolds, 2006; Anderson and Mellor, 2008), although the effect is often marginally significant, not robust to changes in the definition of smoker, or counterintuitive (i.e. smokers being more risk averse) (Reynolds, 2004, 2006). There is, however, an equal (if not larger) number of studies that find no significant association between risk preferences and smoking status (Yi et al., 2007; Reynolds et al., 2007; Harrison et al., 2010; Mitchell, 1999; Galizzi and Miraldo, 2012; Szrek et al., 2012; Harrison et al., 2015a).

Similar mixed results are common for other risky behaviors. Most studies on the external validity of risk preferences measures, moreover, consider student samples and focus on one specific measure of risk preferences and one specific behavior at a time.

We contribute to the existing literature on the external validity of risk preference measures by systematically comparing two experimental measures (B-EG and HL) and three self-reported indicators (SOEP-G, SOEP-F, and SOEP-H) against a comprehensive range of behaviors for a representative sample of the UK adult population. The importance of systematic analyses to assess the external validity of experimental measures is discussed in great detail in Galizzi and Navarro-Martinez (2015) in the context of social preferences games.

3 Experiments

3.1 The Innovation Panel of the UK Household Longitudinal Survey

The artefactual field experiment was conducted using the UK Household Longitudinal Survey (UKHLS, also known as *Understanding Society*), a survey funded by the Economic and Social Research Council (ESRC) and managed by the Institute for Social and Economic Research (ISER) at the University of Essex. The UKHLS is the largest multi-topic household panel survey in the world, interviewing annually more than 40,000 respondents (see `www.understandingsociety.ac.uk`). The UKHLS also includes the Innovation Panel (IP), a parallel longitudinal survey of about 1,500 households, whose questionnaire content mirrors that of the larger survey, but it is explicitly designed "for experimental and methodological research relevant for longitudinal surveys" (Jackle et al., 2015).

Like the UKHLS, the IP has a stratified and geographically clustered sample design (Lynn, 2009). It provides a representative sample of the population of England, Scotland, and Wales, but (unlike the main *Understanding Society* survey) excludes Northern Ireland and Scotland areas north of the Caledonian Canal. Post code sectors from the Postcode Address File were used as primary sampling units (PSUs) and stratified by Government Office Region (GOR). Respondents to the IP routinely take part in randomized controlled experiments. In years 2012 to 2014 (i.e. in IP5 to IP7) the main experimental intervention focused on the use of web interviewing, which has the potential of enhancing response rates and fieldwork efficiency while at the same time reducing survey costs. A detailed description of the experiments conducted in IP5-7 can be found in Jackle et al. (2015). Appendix B describes in greater detail the sampling design of the IP; the sampling weights that are used in all the estimations presented in the rest of the paper; and the random assignment of the households to either face-to-face (F2F) or web (WEB) interview mode.

3.2 Allocation to the risk preferences module

In IP6 a total of $n = 1{,}087$ sampled households were randomly allocated to the risk preferences elicitation module, independently from the (randomly) allocated interview mode. The sample included members of the Wave 4 refreshment sample and original sample members (OSMs) for which a longer longitudinal panel was available. Due to survey non-participation, the final sample of households eligible for the experimental module was $n = 808$, of which 528 (65.3%) randomly allocated to the WEB mode. In each eligible household, one adult respondent was then randomly selected using a Kish grid to answer the risk preferences module. A total of 661 respondents completed the risk preferences module in IP6, and 468 did it one year later, in IP7. A total of 413 individuals answered the risk preferences questions in both IP6 and IP7. Appendix C describes the details of the allocation to the risk preferences module.

The implementation of the risk preferences experimental tests in IP6 was programmed and piloted, in both interview modes, in January 2013. In the same month, professional interviewers from the National Centre for Social Research (NatCen) were selected and trained for the fieldwork. The data collection of the experimental tests in IP6 took place between February and July 2013. The data collection of the experimental tests in IP7 took place between May and November 2014.

3.3 Experimental and survey questions on risk preferences in the UKHLS Innovation Panel

The participants in the experimental module were presented three tasks. The first task was to measure time preferences and is fully described in Galizzi et al. (2016b). The second identified a-temporal risk preferences using both the HL and the B-EG methods. These two tasks took place at the beginning of the individual survey. The third task consisted of an ad hoc questionnaire conducted after all the standard modules of the UKHLS, containing the SOEP survey questions on risk attitudes, and other behavioral dimensions of interest.

3.3.1 Experimental design considerations

In designing the risk preferences questions (and more generally the experimental module) a number of considerations and constraints were taken into account. The most pressing constraint was the desire to keep the overall burden of the experimental module to a minimum in order to minimize respondents' fatigue, non-response, and attrition. Given the available research budget, the entire experimental module was allocated an indicative time slot of about fifteen minutes, which substantially limited the number of questions that could be introduced.[6] Given that the estimation of the various time preferences models required relatively more data points, the module was designed with the idea that about 20 questions were allocated to risk preferences.

The second overall design consideration was the importance of presenting to subjects the different risk tradeoffs in a transparent and easily accessible way. This aspect is of great importance in our representative sample of respondents, given their inherent heterogeneity in socio-demographic conditions, education levels, numeracy, cognitive skills, and familiarity with similar tasks. We thus opted for the HL MPL test, the B-EG ordered lotteries sequence method, and the SOEP risk attitude measures which have been widely adopted in the experimental and behavioral economics literature, and have been successfully employed with general and representative samples of the population in Canada, Denmark, Germany, the Netherlands, and Vietnam, among other countries (Donkers et al., 2001; Andersen et al., 2008b; Dave et al., 2010; Tanaka et al., 2010; Dohmen et al., 2011; Von Gaudecker et al., 2011), among other countries.

[6] There was no binding time limit to answer the experimental questions, though, and subjects were able to answer each question at their own pace. We also have access to meta-data on individual response times for each experimental question, and in fact we use these data in the structural estimations of risk and time preferences in the companion papers (Galizzi et al., 2016b,a).

Risk preferences HL questions consisted of two subsequent blocks of binary choices, the first block of 9 questions referring to relatively low monetary stakes, while the second block of 9 questions presenting options with relatively high monetary outcomes. To maximize clarity and transparency, in each block the list of binary choices was presented in an ordered manner with the probability of the high prize increasing in 0.1 increments (see more below). In both the WEB and F2F response modes each question was presented in a separate screen and the subjects had to confirm their answers before moving to the next screen.

A last question was introduced to measure risk preferences using the B-EG ordered lottery sequence method (see more below). Subjects responded to these risk preferences after having answered the time preferences questions.

To further encourage clarity and transparency of the procedure, and to ensure that subjects understood the consequences of their choices, the risk preferences module started with a preamble which briefly described the tasks, the aim of the experiment, and highlighted that their possible gains were related to their actual choices (see the Appendix D for details).

The third overall design consideration revolved around the exact level of the monetary stakes to be used in the incentive-compatible tests. *Understanding Society* respondents are accustomed to fixed-fees incentives in the region of £5-10. In order to genuinely elicit risk preferences for our heterogeneous set of respondents, we had to make monetary outcomes salient.

Moreover, we faced a trade-off between the simultaneous needs of making the monetary outcomes salient, and diversifying the range of monetary stakes in the attempt to capture possible cases of risk preference depending on the absolute value of the monetary stakes (for instance, being risk averse for low amounts, but risk loving for high amounts, or the other way around: Hey and Orme, 1994; Harrison et al., 2005a; Holt and Laury, 2005; Cox and Sadiraj, 2006; Cox et al., 2015). Due to the binding time constraint, we opted for two MPL series of HL questions with the following prize sets: [A1: £40 and £32; B1: £77 and £2], [A2: £100 and £40; B2: £180 and £2]. That is, a first series of 9 binary questions used relatively low monetary stakes, with a safer lottery giving outcomes of £32-£40, and the other risky lottery returning payments of £2-£77. A second series asked 9 binary questions with relatively high monetary stakes, with a safer lottery giving out payments of £40-£100, and the other risky lottery returning payments of £2-£180.

The multiple-options B-EG test was implemented in a version with a somewhat lower level of monetary stakes: the prize sets were [A1: £28; A2: £28], [B1: £24; B2: £36], [C1: £20; C2: £44], [D1: £16; D2: £52], [E1: £12, E2: £60], [F1: £2; F2: £70].

A fourth, related, overall design consideration concerned the choice of how to administer the experimental tests and to pay out the chosen options. Given the relatively limited budget constraint, the use of salient monetary outcomes did not allow paying out all the experimental tests for all subjects. We thus randomly select one of the experimental tasks for being actually played for real, and for giving each subject one in ten chances to receive the associated payment. Although it is well known that such a procedure transforms the experimental tasks into a compound-lottery (Holt, 1986), it is a common belief among experimental economists that this procedure does not interfere with the genuine elicitation of individual risk preferences.[7]

Another issue was that the payment procedures used to play the experimental tests should be familiar and credible to the respondents. Respondents to the *Understanding Society IP* are accustomed to receive monetary incentives. Credibility of monetary payments was thus not an issue. What respondents were relatively less familiar with was the administration of randomized payments. Randomized payments were administered by means of an *ad hoc* computerized script, which visualized the probability to be among the respondents to be paid, the random selection of the question which determined the payoff function, and the random outcome of the selected lottery. Appendix E describes the randomization procedure. The gains were paid in form of vouchers that could be redeemed both online and at a large number of supermarkets and high street shops all over the UK.

3.3.2 The HL and B-EG tests in detail

In this section we provide a more detailed description of the HL and B-EG task.

Subjects in both the F2F and the WEB mode answered the same questions in the same frame, with a single question presented at a time in a separate screen. Questions in the first block of 9 HL questions with relatively low monetary stakes (*low*) were presented first, followed by questions in a second block of 9 HL questions with relatively high monetary payments (*high*). Within each block of questions, choices were presented in an ordered manner, with the probability of the high prize increasing in 0.1 increments. Unlike Hey and Orme (1994) we did not use any pie chart showing prizes and probabilities. The typical wording of the risk preferences questions was as follows:

> Between Lottery A and Lottery B, which lottery do you prefer?
> Options:
> LOTTERY A: Winning £40 with a 1 in 10 chance (i.e., die roll is 1), and £32 with a 9 in 10 chance (i.e., die roll is 2,3,4,5,6,7,8,9 or 10)
> LOTTERY B: Winning £77 with a 1 in 10 chance (i.e., die roll is 1), and £2 with a 9 in 10 chance (i.e., die roll is 2,3,4,5,6,7,8,9 or 10)

Table 2 shows the specific values of the lottery stakes and probabilities from the HL questions.

The B-EG task required the subjects choosing between 6 lotteries: A, B, C, D, E and F. Each lottery gave a 50% chance of receiving a low cash payment and a 50% chance of a high cash payment. The payments for the lotteries were: A: low = £28, high = £28; B: low = £24, high = £36; C: low = £20, high = £44; D: low = £16, high = £52; E: low = £12, high = £60; F: low = £2, high = £70. Table 3 shows the B-EG lotteries. These choices were thus increasing in the variance of the outcomes and in the risk they represented, with

[7] See for instance Conlisk (1989); Starmer and Sugden (1991); Wilcox (1993); Bernasconi (1994); Beattie and Loomes (1997); Cubitt et al. (1998); Harrison et al. (2002); Drehmann et al. (2005); Harrison et al. (2007b); Harrison and Rutström (2008). See Baltussen et al. (2012), Harrison and Swarthout (2014), Cox et al. (2015), Harrison et al. (2015b), and March et al. (2016) for some recent studies looking more closely at the random lottery incentive mechanism.

A being the safe choice (a guaranteed payment of £28 and thus a variance of $\sigma_A^2 = 0$) and F the highest-risk choice (a variance of $\sigma_F^2 = 1156$). To make a choice, a subject chooses one out of the six lotteries. The B-EG measure thus increases with an individual's appetite for risk. It is worthwhile to notice that any subject risk neutral or risk averse should always prefer lottery E to lottery F. This is because, by construction, lottery E has the same expected value of lottery F (equal to £36 in both cases) but smaller variance.

3.3.3 Questionnaire

The questionnaire included survey questions on self-reported risk attitudes and other related behavioral attitudes of interest. We included a question about self-reported attitude towards risk-taking in general (Dohmen et al., 2011), taking values from 0 ("I am generally a person unwilling to take risks") to 10 ("I am generally a person fully prepared to take risks") (SOEP-G); a question about self-reported attitude towards risk-taking in health, taking values from 0 ("I am generally a person unwilling to take risks in health") to 10 ("I am generally a person fully prepared to take risks in health") (SOEP-H); as well as a question about self-reported attitude towards risk-taking in financial matters, taking values from 0 ("I am generally a person unwilling to take risks in financial matters") to 10 ("I am generally a person fully prepared to take risks in financial matters") (SOEP-F).[8]

In addition to these variables from the questionnaire, in our analysis we use a number of variables from the main IP survey. In the analysis reported here we use individual and household characteristics of the respondents such as: household per-capita income (*PCIncome*); family size (*FamSize*); respondents' age (*Age*), gender (*Female*), and marital status (*Married*); dummy variables for whether the respondents are employees (*Empl*), self-employees (*SelfEmpl*) or unemployed (*Unempl*); for whether they have achieved at least an upper secondary school degree (*ALevel*); for whether they report to smoke (*Smoker*), drink more than 5 alcohol units in the last four weeks (*Heavy Drinking*); for whether the respondents answer the survey in the WEB mode (*Web*).

We also use some wave-specific variables that are only available in either IP6 or IP7. For instance, in IP6 the height and weight of the respondents were directly measured, which allows us to calculate the subjects' body mass index (*BMI*). In IP6, moreover, there are variables for the weekly consumption of junk food (*Junk Food*), as well as dummy variables for whether the subjects are regular savers (*RegularSaver*), have a private pension fund (*PrivatePension*), and a categorical variable for the allocation of the savings between short and long term (*Savings Horizon*).

In IP7 there is a variable for the weekly consumption of fruit and vegetables (*Fruit and Vegetables*), and we have also included a simplified version of the DOSPERT test for risk-taking in hypothetical domain-specific situations. The DOSPERT questions ask respondents to self-report how likely they are to engage in a series of specific activities or behaviors, using a 7-points Likert scale, taking value from 1 ("Extremely unlikely") to 7 ("Extremely likely"). In particular, to ensure comparability with the SOEP-F and the SOEP-H questions we have introduced six finance-related and six health-related hypothetical questions. They ask IP7 respondents what is the likelihood of: betting a day's

[8]The questions are shown in Appendix F.

income at the horse race (*Bet: horses*); investing 10% of their annual income in a moderate growth mutual fund (*Invest 10%: mutual fund*); betting a day's income at a high-stake poker game (*Bet: poker game*); investing 5% of their annual income in a very speculative stock (*Invest 5%: speculation*); betting a day's income on the outcome of a sporting event (*Bet: sports*); investing 10% of their annual income in a new business venture (*Invest 10%: business*); drinking heavily at a social function (*Drink heavily: social*); engaging in unprotected sex (*Unprotected sex*); driving a car without wearing a seat belt (*Drive w/o seatbelt*); riding a motorcycle without a helmet (*Motorcycle w/o helmet*); sunbathing without sunscreen (*No sunscreen*); and walking home alone at night in an unsafe area of the town (*Walk home alone unsafe*). Following the literature (Weber et al 2002, Blais & Weber, 2006), we aggregate the responses to the DOSPERT questions in two summary indexes for risk taking in health and in finance. The full set of questions is available in Appendix F. Some descriptive statistics for the main variables above in our sample are reported in Section 5.1 and Table 4.

4 Theoretical framework and empirical analysis

In Section 5.1 we first present results in terms of raw choices in the different tasks. In particular, in the B-EG task, we report the proportions of subjects who chose the different lotteries, from lottery A to lottery F. Similarly, in the HL we report the proportions of subjects who switch for the first time from lottery A to lottery B in correspondence of each HL question in the sequence of pairwise choices, from 1 (choosing for the first time the riskier lottery B when $p = 0.1$) to 2 (choosing the riskier lottery B when $p = 0.2$) and so forth, up to 10 (never choosing the riskier lottery B). In the SOEP task, the self-reported attitude towards risk-taking takes values from 0 ("I am generally a person unwilling to take risks") to 10 ("I am generally a person fully prepared to take risks").

The different tasks point to a subject's appetite for risk into different directions. The higher is the first switching point in the HL task, in fact, the higher is the subject's risk aversion. The higher is the SOEP score, however, the higher is the subject's risk taking attitude. Similarly, in the B-EG task the subject is less risk averse when moves from lottery A (the certain gamble) to lottery F (the riskiest gamble). Thus, while the risk aversion increases with the raw choices in HL task, risk aversion decreases with the lottery choices in the B-EG task, and risk taking attitude increases with the Likert scale point in the SOEP tasks.

In order to compare the three measures in a meaningful and intuitive way, we therefore reverse-code the responses to the B-EG and the SOEP tasks, so that all the three tasks increase with risk aversion and decrease with risk taking. In Sections 5.3, 5.4, and 5.5 we present the pairwise correlations, and the statistics about the cross-validity and the temporal stability of the B-EG, HL, and SOEP tasks, respectively, in terms of these recoded measures increasing with risk aversion and decreasing with risk taking. In particular, we define HL RA as an increasing measure of risk aversion according to the HL task, that is increasing with the point in which the subject switches for the first time from lottery A to lottery B in the sequence of the HL questions, taking values from 1 (choosing for the first time the riskier lottery B when $p = 0.1$) to 2 (choosing the riskier lottery B when $p = 0.2$) and so forth, up to 10 (never choosing the riskier lottery B). We define B-EG RA as an increasing measure of risk aversion according to the B-EG task, taking values from 1 (very risk seeking) if the subject chooses lottery F (the riskiest gamble) in the B-EG task, to 6 (very risk averse) if the subject chooses lottery A (the certain gamble) in the B-EG task. Similarly, we define SOEP-G RA as a decreasing measure of risk taking attitude according to the SOEP-G task, taking value from 1 (willing to take risk) if the subject chooses to 10 in the SOEP-G Likert scale ("I am generally a person fully prepared to take risks"), to value 10 (unwilling to take risks) if the subject chooses 0 in the SOEP-G Likert scale ("I am generally a person unwilling to take risks"). We equivalently define SOEP-F RA and SOEP-H RA with the corresponding meaning of risk taking attitudes in finance and in health, respectively. It is expected therefore that the pairwise correlations between the B-EG RA, HL RA and SOEP RA measures are positive, and that the three measures cross-validate each other in the same direction.

Notice that focusing on correlations without a detailed cross-tabulation of the choices can be misleading. For example, correlations could not capture the fact that, regardless of what they choose in the B-EG task, many subjects in the HL task either switch immediately to lottery B or never switch to lottery B. There is, moreover, a key issue related to expressing risk aversion simply as a mere linear function of the first switching point in the series of the HL questions or of the lottery choice in the B-EG task. The problem is, clearly, that subjects' risk aversion is a *non-linear* function of the switching points in the HL questions and of the lottery choice in the B-EG task. We need therefore to map the choices in the HL and the B-EG tasks into risk aversion parameters within a theoretical framework, like in Loomes and Pogrebna (2014) and Crosetto and Filippin (2015).

In the rest of the analysis, we assume that EUT holds for choices over risky options.[9] We imagine that subjects use some utility function $U(\cdot)$ that is separable and stationary over time. In particular, following a broad economics literature (e.g Wakker, 2008), we assume that subjects use a CRRA utility function

$$U(M_\tau, W) = \frac{(W + M_\tau)^{(1-r)}}{1 - r} \tag{1}$$

for $r \neq 1$, where r is the CRRA coefficient, and W is the individual level of background income within which the monetary prize M is integrated. For $r = 1$, $U(M_\tau) = \ln(W + M_\tau)$ can be assumed if needed. Within this CRRA functional form, $r = 0$ implies risk-neutral choices, $r > 0$ denotes risk aversion, and $r < 0$ implies risk-seeking behavior. Here we assume that $W = 0$ for all subjects. The role of heterogeneous individual-specific levels of background income and of its possible integration with the B-EG and HL experimental prizes on the estimation of risk preferences is explored in a companion paper (Galizzi et al., 2016a).

[9]We do not consider here the question of whether EUT is the most accurate model for risk preferences. Rather, this standard model of risk preferences can be viewed descriptively as a convenient statistics to study the validity and stability of individual responses to different risk preferences measures. Our data, in fact, allows us to estimate also non-EUT models of risk preferences such as the Rank Dependent Utility model by Quiggin (1982), a possibility that it is explored by Galizzi et al. (2016b).

Assuming EUT and CRRA (and assuming that the experimental prizes are integrated in a background income equal to zero) the coefficient of relative risk aversion associated to each switching point in the HL task is described by the intervals reported in the right-most column in Table 2. As it can be seen the relation between coefficients of relative risk aversion and switching points in the HL task is indeed non-linear.

Similarly, under the same assumptions, the coefficient of relative risk aversion associated to each lottery choice in the B-EG task is described by the intervals reported in the right-most column of Table 3. Also in this case the relation between the coefficients of relative risk aversion and the lottery choices in the B-EG task is non-linear.

To associate the observed lottery choices to an appropriate level of relative risk aversion, we run a set of interval regressions for the B-EG and the HL tasks using as intervals the lower and upper bounds of relative risk aversion $[r_l, r_u]$ consistent with the EUT and CRRA assumptions reported in Tables 2 and 3. This allows us to provide an estimate of $E[r^*|r_l < r^* \leq r_u; X]$, where r^* is the latent relative risk aversion, and X is a set of observed characteristics of the respondents, and r_l and r_u are the lower and upper bound consistent with the actual choices in the B-EG and HL tasks (Section 5.4.4). We use these estimated values in the assessment of the cross-validity, temporal stability, and external validity of the risk preferences measures (Sections 5.4, 5.5, and 5.6, respectively).

In Section 5.6 we conduct a systematic external validity analysis to predict a range of health and financial behaviours using as predictors the SOEP measures of risk attitude, the CRRA values conditional to the specific lottery choice in the B-EG task estimated by the above interval regressions, and the CRRA values conditional to the specific switching point in the HL task estimated by the above interval regressions.

All the cross-validity, temporal stability, and external validity regressions have been replicated using a comprehensive set of alternative models and specifications using either the raw lottery choices and switching points, or the midpoints of the associated intervals of the CRRA, or the estimated values described above.

In particular, we have associated each interval of CRRA values associated to a switching point in HL - or a lottery choice in B-EG - to the midpoint of that CRRA interval. The midpoints can be easily calculated for the closed-end CRRA ranges corresponding to the switching points 2 through 9 in the HL tasks and to the lottery choices B to E in the B-EG task. Fitted values can be estimated for the midpoints of the open-ended intervals in the B-EG and HL tasks.

All estimations allow for the complex survey design, use appropriate sampling weights to draw valid inference for the adult population in the UK, and adjust standard errors at strata and PSU levels. In all the cross-validity, temporal stability, and external validity regressions, we have found the same patterns of significant associations across these sets of regressions using the raw responses, the midpoints of the CRRA intervals, and the estimated CRRA values described above. We therefore report the results for the interval regressions estimations, while the results of all further cross-validity, temporal stability, and external validity regressions using the raw responses and the CRRA midpoints are available on request.

5 Results

We first present some simple descriptive statistics for the respondents in our UK representative sample (5.1). We then describe the subjects' responses to the B-EG, HL, and the SOEP tasks in IP6, in IP7, and by comparing responses in IP6 and IP7 (5.2). We next present some pairwise correlations among the B-EG and HL measures of risk aversion, and the SOEP risk attitudes measures (5.3). We then directly deal with each of the three issues of cross-validity (5.4), temporal stability (5.5), and external validity (5.6) of the three risk preferences measures.

Although 808 households in the IP were randomly assigned to the experimental module, only 661 eligible respondents actually answered the risk and time preferences questions, and they could not be representative of the UK population. Therefore, in order to adjust for differential attrition and non-response and to generalize results to the UK adult population, all the statistics and the estimations allow for the complex survey design using the appropriate sampling weights to draw valid inference for the adult population in the UK, and adjust standard errors at strata and PSU levels. Moreover, when we use the panel component to compare responses across the two waves and to look at the temporal stability, we use longitudinal sampling weights that adjust for attrition.

5.1 Descriptive statistics

The main characteristics of our sample in IP6 are reported in Table 4 in Appendix A. The top half of the table reports the descriptive statistics of all the respondents who answered the risk preferences questions in IP6. Using sampling weights reduces the number of observations because in IP6 and IP7 some respondents received a zero cross-sectional weight, namely the respondents in the households containing no original sample members with a non-zero longitudinal weight (Jackle et al., 2015). The bottom half reports the descriptive statistics only for the respondents in IP6 who answered the risk preferences questions in both IP6 and IP7.

As it can be seen, 52.8% percent of all the IP6 respondents were female, and 43.6% were married. The average age was 49.47 years old. About 41.7% of the respondents had at least an A-level education qualification (that is, have obtained at least an upper secondary school degree). The average level of gross household per capita monthly income was of £1,966. About 3.9% of respondents were unemployed and around 33.1% participate to the survey via WEB.

The subjects who answered the risk preferences questions in both IP6 and IP7 had substantially similar characteristics to the ones described above, with the main differences being that, on average, they were younger (47.53% years old, on average), had higher monthly income (£2,124), were more likely to be married (53%) or employees (53% compared to 45.6% in IP6), to have an A-level qualification (47.6%), or to respond to the survey via WEB (38.2%).

5.2 Responses

5.2.1 Responses in IP6

Figure 1 reports the histograms of the raw responses in the HL, B-EG and SOEP tasks.

Concerning the HL task, it is generally expected that respondents should be consistent with their choices. In other words, once in the sequence of HL questions they have chosen lottery B, they should not choose lottery A in any of the following questions, where the expected value of the lottery B further increases, that is, they should not "switch back" to A.

A common finding in the lab and field experiments that do not force consistency in individual responses, is that a conspicuous proportion of subjects make "inconsistent" sequences of binary choices in the sense that, after having chosen lottery B, they "switch back" to lottery A. Crosetto and Filippin (2015) present a meta-analysis of 30 published experiments conducted using the HL method, involving 4,726 participants, mostly university students. They show that, on average, about 16.3% of the experimental subjects make inconsistent choices, with the proportion being higher in the few experiments conducted with more general samples of the population.[10] Responses in our sample are not an exception to this general finding.

First, a total of 657 subjects started answering the risk preferences questions. Because two subjects did not complete the HL questions with low monetary stakes and another subject did not complete the HL questions with high monetary stakes, a total of 655 and 654 subjects answered the HL questions with low and high monetary stakes, respectively. Among the 655 subjects who answered the HL questions with low monetary stakes, 160 subjects (24.43% of the sample) made at least one inconsistent choice when answering HL questions with low monetary stakes. Among the 654 subjects who started answering HL questions with high monetary stakes, 126 subjects (19.26% of the sample) made at least one inconsistent choice when answering HL questions with high monetary stakes. There is a core group of 80 respondents who made at least one inconsistent choice in both the low and the high monetary stakes HL questions.

In what follows, we separately present results for the whole sample of respondents, and focusing only on the group of the "consistent" subjects who never behaved inconsistently in the HL questions in IP6. There is a total of 206 subjects who behaved inconsistently at least once in IP6, in the sense that they switched back to lottery A after choosing lottery B at least once in the HL task. In Section 6 we look more closely to the subjects who behaved only "locally" inconsistently in the sense that they made only one "mistake" in the sequence of HL choices, but their choices were otherwise consistent. In Section 6 we pool these locally inconsistent subjects together with the consistent subjects to see whether it makes any difference in the estimations. It turns out that it does not.

Table 5 reports the number of subjects who chose lottery B for the first time in correspondence of each HL question, also splitting the sample into consistent and inconsistent subjects in the HL tasks with low and high monetary stakes in IP6 and IP7.

[10] Charness and Viceisza (2015) review the studies using the HL task in developing countries and find higher proportions of inconsistent subjects, between 14% and 66.5% of the subjects.

Furthermore, by looking at the point in the sequence where the respondent switches from lottery A to lottery B, one can calculate the lower and upper bounds of the underlying relative risk aversion parameter r under the EUT and the CRRA assumptions (Section 4).

When looking at the number of safe choices in the HL task for all subjects together, including the inconsistent (Table 6, All subjects), it can be seen that about a fifth of the respondents never chose the safe lottery A: 17.71% always chose lottery B with low monetary stakes, while 23.39% always chose lottery B with high stakes. On the other hand, a roughly equal proportion of subjects always chose the safer lottery A in all series of choices (21.07% and 19.42% with low and high monetary stakes, respectively). Apart from these two polar choices, the distributions of responses to the HL tasks is uni-modal and roughly symmetric, with the bulk of subjects choosing lottery A between three and six times in the series of choices.

The responses are more polarized when looking at the 495 and 528 subjects who never switched back in the HL questions with low and high stakes, respectively, thus excluding the 160 and 126 inconsistent respondents, respectively (Table 6, Consistent). Between a fourth and a third of these respondents never chose the safer lottery A (23.43% and 28.98% with low and high monetary stakes, respectively).

In IP6 our sample of respondents provided heterogeneous responses also in the B-EG task (Table 7). More than a third of the 655 subjects who answered the B-EG task (n=230, 35.11% of the respondents) chose lottery A, the certain gamble, in the B-EG task. The riskiest gamble, lottery F, was chosen by 104 subjects, 15.88% of the total. A total of 120 and 84 respondents chose lotteries C and D, respectively (18.32% and 12.82% of the total, respectively). Lotteries B and E were chosen by 64 and 53 subjects, respectively (9.77% and 8.09% respectively).

Table 8 reports the responses to the SOEP tasks for the 629 subjects who answered the task in IP6. The mode of the distribution of the responses has a willingness to take risk equal to 5, in general, and equal to 2 in finance and health. The average willingness to take risk was 4.52 in SOEP-G, 3.17 in SOEP-F, and 3.15 in SOEP-H. Appendix G reports analogous information for the responses in IP7, and for the subsample of subjects who participated to the risk preferences module in both IP6 and IP7.

5.3 Correlations

In preparation of our discussion on the cross-validity of the three risk preferences measures at an individual level, we present the aggregate figures for pairwise correlations between the B-EG and HL measures of risk aversion and the SOEP risk attitudes measures. As described above, B-EG RA is an increasing measure of risk aversion according to the B-EG task, taking value 1 (very risk seeking) if the subject chooses lottery F (the riskiest gamble) in the B-EG task, to value 6 (very risk averse) if the subject chooses lottery A (the certain gamble) in the B-EG task. Similarly, HL RA is an increasing measure of risk aversion according to the HL task, in terms of the point in which the subject switches for the first time from lottery A to lottery B in the sequence of HL questions, from 1 (choosing for the first time the riskier lottery B when $p = 0.1$) to 10 (never choosing the riskier lottery B). Finally the SOEP-G RA is a decreasing measure of risk taking attitude according to the SOEP-G task, taking value 1 (very willing to take risks) if the subject chooses 10 in the SOEP-G Likert scale ("I am generally a person fully prepared to take risks"), to 10 (unwilling to take risks) if the subject chooses 0 in the SOEP-G Likert scale ("I am generally a person unwilling to take risks"), with the equivalent meaning for SOEP-F and SOEP-H. It is expected therefore that the pairwise correlations between the B-EG RA, HL RA and SOEP RA measures are all positive.

Table 9 reports the Pearson cross-methods pairwise correlations with sampling weights between the different measures of risk aversion, together with their significance levels, at IP6 (top half) and IP7 (bottom half).[11] In both waves the HL measures of risk aversion with low monetary stakes are strongly and significantly positively correlated with the HL measure of risk aversion with high monetary stakes: the correlation is 0.669 ($p < 0.001$) in IP6, and 0.627 ($p < 0.001$) in IP7.

The HL measures of risk aversion, with both low and high stakes, are also significantly correlated with the B-EG measure of risk aversion, although the correlation is generally low (0.120, $p < 0.001$, and 0.129, $p < 0.001$, between responses in B-EG and HL choices with low stakes in IP6 and IP7, respectively; 0.172, $p < 0.001$, and 0.169, $p < 0.001$, between responses in B-EG and HL choices with high stakes in IP6 and IP7, respectively). The finding of a significant but low correlation between B-EG and HL measures in our UK representative sample is generally in line with the evidence of no to low correlations from the previous literature (Dave et al., 2010; Reynaud and Couture, 2012; Deck et al., 2013; Loomes and Pogrebna, 2014).[12]

In both IP6 and IP7, responses to the three self-reported SOEP risk attitude measures are highly and significantly correlated among them: 0.621 and 0.4989 between SOEP-G and SOEP-F in IP6 and IP7, respectively; 0.531 and 0.420 between SOEP-G and SOEP-H in IP6 and IP7, respectively; and 0.565 and 0.439 between SOEP-F and SOEP-H in IP6 and IP7, respectively (all with $p < 0.001$). These highly significant correlations between the various SOEP measures are in line with what is typically found in the literature (Dohmen et al., 2011; Falk et al., 2016; Vieider et al., 2015a; Josef et al., 2016): Vieider et al. (2015a), for instance, find correlations of 0.495 between SOEP-G and SOEP-F, 0.137 between SOEP-G and SOEP-H, and 0.250 between SOEP-F and SOEP-H.

In IP6 the B-EG measure of risk aversion is significantly and positively correlated with the SOEP risk attitude measures but the correlations are low: 0.095 (p =0.058) with the SOEP-G, 0.136 (p =0.007) with the SOEP-F, and 0.103 (p =0.015) with the SOEP-H. In IP7 the B-EG measure of risk aversion is not significantly correlated with any of the SOEP measures: all correlations are close to zero. This no to low correlation between

[11]We have also calculated Spearman correlations and polychoric correlations to account for the ordinal structure of the data. All correlations are substantially similar and are available on request.

[12]It is also worthwhile to note that the low correlations between the B-EG and the HL measures of risk aversion in a within-subject experiment like ours could, in principle, be partly explained by some form of "hedging" across incentive-compatible tasks attempted by risk-seeking subjects, a possibility noticed by Crosetto and Filippin (2015) who, for this reason, adopt a between-subject design in their experiment. In our specific experimental design, however, the scope for such "hedging" strategies is realistically quite limited because subjects knew in advance that either the B-EG question or one of the HL questions would be randomly selected for real payments.

the B-EG and the SOEP measures in our UK representative sample is different from a significant correlation of around 0.30 found by Crosetto and Filippin (2015) in a sample of students in Germany.

In IP6 there is no significant correlation between SOEP-H and SOEP-F, on the one hand, and either HL task, on the other hand. There is also no significant correlation between the SOEP-G and the HL measures with low monetary stakes. There is positive significant but very low correlation between the SOEP-G and the HL measures with high monetary stakes: 0.014, p =0.013. In IP7 there is virtually no significant correlation between SOEP and HL measures: four out of six pairwise correlations are negative, including a marginally significant correlation between the HL measure with high monetary stakes and the SOEP-H measure (-0.125, p =0.052). This substantial lack of correlation between the HL and the SOEP measures in our UK representative sample is in contrast with the small but significant associations found by Dohmen et al. (2011) and Josef et al. (2016) in representative samples of the German population, and with the associations found by Attanasi et al. (2016) and Falk et al. (2016) in student samples in Italy and Germany, respectively, but is consistent with the lack of correlation found by Vieider et al. (2015a) in a student sample in the UK (as well as in other 13 out of 30 countries)), and by Szrek et al. (2012) and Charness and Viceisza (2015) in convenience samples in South Africa and Senegal, respectively. It is also generally in line with the weak and marginally significant correlations also found by Wölbert and Riedl (2013) and Crosetto and Filippin (2015) in student samples in the Netherlands and Germany, respectively.

Table 10 reports the cross-waves pairwise correlations, together with their significance levels, for the B-EG and HL measures (top half) and the SOEP measures (bottom half) for the subjects for which we have responses in both waves. Table 10 confirms that the main results discussed above hold for the subset of subjects for whom B-EG, HL, and SOEP measures can be linked across waves, namely: that the choices are significantly and highly positively correlated among HL questions with low and high monetary stakes (0.694 and 0.609 in IP6 and IP7, respectively, both with $p < 0.001$); and that the responses are significantly positive, but moderately, correlated between the B-EG and the HL task (for the HL task with low monetary stakes, 0.061, $p = 0.357$, and 0.196, $p = 0.009$, in IP6 and IP7, respectively; for the HL task with high monetary stakes, 0.131, $p = 0.01$, and 0.139, $p = 0.034$, in IP6 and IP7, respectively).

Moreover, there are significant and positive correlations of the B-EG and HL measures of risk aversion and between the SOEP measures of risk attitudes across the two waves: the IP6-IP7 correlation is 0.138 for B-EG RA (p =0.028), 0.225 for HL RA with low stakes (p =0.003), 0.219 for HL RA with high stakes, 0.551 for the SOEP-G RA (p <0.001), 0.473 for the SOEP-F RA (p <0.001), and 0.447 for SOEP-H RA (p <0.001). The evidence of significant positive correlation over time for the HL measures is generally in line with the findings of Harrison et al. (2005b) and Andersen et al. (2008b). The evidence of significant correlation of the SOEP measures over time in our UK representative sample is closely in line with what found in Germany by Josef et al. (2016) using repeated waves of the GSOEP: for example, Josef et al. (2016) find correlations in the 0.45-0.53 range for the SOEP-G, in the 0.42-0.50 range for the SOEP-F, and in the 0.38-0.45 range for the SOEP-H.

5.4 Cross validity

We now look at individual responses at one point in time (IP6) and explore how the B-EG and HL measures for risk aversion and the SOEP risk attitudes measures relate to each other. We start presenting the main within-subjects raw responses across each pair of risk preferences tasks: between the B-EG and the SOEP tasks (5.4.1); between the HL and the SOEP tasks (5.4.2); and between the B-EG and the HL tasks (5.4.3). We then estimate the CRRA values associated to each specific lottery choice in the B-EG and to each specific switching point in the HL task, using interval regressions (5.4.4). We next conduct a set of cross-validity ordered probit regressions, where the dependent variables are the SOEP measures of risk attitudes and the main explanatory variables are the CRRA values for the B-EG and HL tasks, estimated by the above interval regressions (5.4.5). We finally run a set of cross-validity interval regressions where the dependent variable is the CRRA value conditional to the lottery choice in the B-EG task - estimated by interval regressions - as a function of the SOEP measure of risk attitudes, and of the CRRA values conditional to the specific switching points in the HL task - also estimated by interval regressions - or where the dependent variable is the CRRA value conditional to the switching point in the HL task - estimated by interval regressions - as a function of the SOEP measure of risk attitudes, and of the CRRA values conditional to the specific lottery choice in the B-EG task - also estimated by interval regressions (5.4.6).

5.4.1 Responses across B-EG and SOEP tasks

Responses to the B-EG and the SOEP task can be directly compared for all respondents who answered both tasks either in IP6 or in IP7. In IP6 a total of 600, 599, and 600 subjects answered both the B-EG and the SOEP-G, SOEP-F, and SOEP-H questions, respectively. A total of 429 subjects answered the B-EG, SOEP-G, and SOEP-F in IP7, and 430 subjects answered the B-EG and the SOEP-H in IP7. Tables 17-19 report the whole distribution of the responses to the B-EG task and to the SOEP-G, SOEP-F, and SOEP-H questions in IP6 and in IP7. For each value of the self-reported SOEP measure of risk attitudes (the rows of the tables), each cell contains the proportion of subjects in each range of CRRA values associated with the lottery choice in the B-EG task (the columns of the tables). If there is stability of risk preferences in the cross-validity sense, one would expect that the subjects who in the SOEP tasks (and especially in the SOEP-F question) self-report to be unwilling to take risks are also the subjects most likely to pick the safer lotteries in the B-EG task (e.g. a CRRA between 3.46 and $+\infty$, corresponding to the safe lottery).

As it can be seen from Table 18, in IP6 there is an association between lower levels of self-reported willingness to take risk in finance (SOEP-F) and safer lottery choices in the B-EG task. For instance, 46.63% who self-reported to be unwilling to take risks in financial matters picked lottery A as their favorite lottery in the B-EG task. In IP6, however, the association of responses between the SOEP-F and the B-EG tasks is far from being perfect. For instance, there are numerous subjects who reported little willingness to take financial risks in the SOEP-F question but then picked risky lotteries in the B-EG task even when they had the option of choosing a safe lottery. More generally, about two thirds of the subjects who reported low levels of willingness to take financial risks in the SOEP-F task, also chose lotteries involving some degree of risk. Also in IP7 there is some general but imperfect pattern of association between lower levels of the willingness to take risk in the SOEP-F question and the choice of safer lotteries in the B-EG task (Table 18).

Table 17 shows that there is also an imperfect association of responses between the SOEP-G and the B-EG tasks, both in IP6 and IP7. The imperfect pattern of associations with the B-EG measure of risk aversion is particularly evident in the responses to the SOEP-H attitudes question (Table 19).

All in all, the analysis of the within-subject individual responses confirms that the individual responses show some general patterns of associations between the B-EG task and the SOEP questions, but those patterns are far from being systematic or perfect, especially for the willingness to take risks in general and in health. It is also worthwhile to note that, conceptually, it is difficult to ascertain how much coherence one should expect to see between the two different measures, given that it is not possible to associate a specific set of CRRA ranges to the different choices in the SOEP questions.

5.4.2 Responses across HL and SOEP tasks

Responses to the SOEP questions can also be compared to responses to the HL tasks for all respondents who answered both tasks either in IP6 or in IP7. In IP6 a total of 600, 599, and 600 subjects answered both the HL task with low and high monetary stakes and to the SOEP-G, SOEP-F, and SOEP-H questions, respectively. In IP7 a total of 431, 430, and 430 subjects answered both the HL tasks (with low and high monetary stakes) and to the SOEP-G, SOEP-F, and SOEP-H questions, respectively.[13] Tables 20-22 report the whole distribution of the responses to the HL measure of risk aversion with low and high monetary stakes and to the SOEP-G, SOEP-F, and SOEP-H risk attitude in IP6. Tables 23-25 report the same distributions for the consistent subjects, who never "switch back" in IP6. For each value of the self-reported willingness to take risk according to the SOEP measure of risk attitudes (the rows of the tables), each cell contains the proportion of subjects in each range of CRRA values associated to the first switching point in the HL task (the columns of the table). If there is stability of risk preferences in the cross-validity sense, one would expect that the subjects who in the SOEP tasks (and especially in the SOEP-F question) self-report to be unwilling to take risks are also the subjects who switch later to the riskier lottery B in the HL task (i.e. a very risk averse subject should switch to lottery B when p is very high, or should never switch to lottery B).

As it can be seen from Table 21, in IP6 there is some association between lower levels of self-reported willingness to take risk in finance (SOEP-F) and the range of CRRA values associated with the first switching point in the HL tasks. For instance, 28.06% and 21.90% of the respondents who self-reported to be *unwilling to take risks in financial matters* never switched to lottery B in the HL tasks with low and high monetary stakes, respectively. In IP6, however, the association of responses between the SOEP-F and the HL tasks is not perfect. More generally, around one half of the subjects who reported low

[13]Results available upon request.

willingness to take financial risks in the SOEP-F task, also chose to switch to lottery B at a point that reveals some degree of risk-seeking in the HL tasks. Also in IP7 there is some but imperfect association between the SOEP-F and the HL tasks (table available upon request). Patterns of weak association between the CRRA ranges associated with the switching points in the HL tasks and the self-reported risk attitudes are also observed in the responses to the SOEP-G and SOEP-H tasks, in both IP6 (Table 20 and 22) and IP7 (table available upon request).

The imperfect association between self-reported willingness to take financial risks and the CRRA ranges associated with the switching points in the HL tasks is also observed when the analysis focuses on the subsample of the 412 subjects who, in IP6, never switched back in the HL tasks (Tables 23-25). For the inconsistent subjects there is essentially no meaningful association between the CRRA ranges associated with the switching points in the HL tasks and the SOEP-F or SOEP-H (tables also available on request).

In sum, the analysis of the within-subject individual responses confirms what was already suggested by the weak correlations between the HL and the SOEP measures in our UK representative sample: the individual responses show some patterns of associations between the HL and the SOEP tasks, but those patterns are far from being systematic or perfect, especially for the risk attitudes in general and in health.

5.4.3 Responses across B-EG and HL tasks

Responses to the HL tasks can be finally compared to responses in the B-EG task for all respondents who answered both tasks in IP6 or IP7. In IP6 a total of 655 and 654 subjects answered both the B-EG task and the HL task with low and high monetary stakes, respectively. In IP7 a total of 449 and 448 subjects answered the B-EG task and the HL tasks with low and high monetary stakes, respectively (tables available upon request). Table 26 reports the whole distribution of the B-EG and HL with low and high monetary stakes measures of risk aversion in IP6. Table 27 reports the same distribution for the consistent subjects in IP6. For each value of the B-EG measure of risk aversion (the rows of the tables), each cell contains the proportion of subjects in each range of the CRRA values associated with the first switching point (the columns of the tables). If there is stability of risk preferences in the cross-validity sense, one would expect that the subjects who in the B-EG task chose the safer lotteries, and in particular the riskless lottery A (a CRRA between 3.46 and $+\infty$), are also the subjects who switch later to the riskier lottery B in the HL task (i.e. a very risk averse subject should switch to lottery B when p is very high, or should never switch to lottery B).

In IP6 there is some but imperfect association between choosing safer lotteries in the B-EG task and the range of CRRA values associated with the first switching point in the HL tasks. For instance, 28.66% and 28.57% of the 230 respondents who chose the riskless lottery A in the B-EG task never switched to lottery B in the HL tasks with low and high monetary stakes, respectively. There are however also as many as 32.98% and 33.26% of the subjects who chose the riskless lottery A in the B-EG task but then switched to the risky lottery B immediately in the HL sequence of pairwise choices with low and high monetary stakes, respectively. Similarly imperfect patterns of associations of responses to the B-EG and HL tasks can be observed in IP7 (table available upon request) or focusing on the consistent subjects who never switch back (Table 27). In sum, there is some pattern of association between the responses to the B-EG and the HL tasks, but these patterns are neither perfect nor systematic.

5.4.4 Estimated CRRA values for the B-EG and HL measures of risk aversion

As discussed above, there is a key issue related to expressing risk aversion simply as a mere linear function of the first switching point in the series of the HL questions or of the reverse-coded lottery choice in the B-EG task. The problem is that subjects' risk preferences are *non-linear* functions of the switching points in the HL questions and of the lottery choice in the B-EG task.

We run a set of interval regressions for the B-EG and the HL tasks, where the dependent variables are the lower and upper bounds of the CRRA values associated to each switching point in the HL task and to each lottery choice in the B-EG task. We first run simple interval regressions for the B-EG task, and for the HL tasks with low and high monetary payoffs and we report the estimated CRRA values for each task in IP6 for all the subjects and for the consistent subjects only (Table 28). This allows a direct comparison of the estimated CRRA across the B-EG and the HL tasks for our UK representative sample.

We then replicate the same interval regressions for each task adding a list of control variables (Table 29). This allows to directly investigate the association of each observable characteristic with the estimated CRRA, and to directly compare their sign and magnitude across the different tasks.

We next estimate the CRRA values conditional to each specific switching point in the HL task and to each specific lottery choice in the B-EG task, and we use these estimates for the B-EG and the HL measures of risk aversion in the cross-validity analysis in Sections 5.4.5 and 5.4.6. This allows us to replicate the cross-validity analysis in an ordered probit or an interval regression framework, and to systematically look at the way the three measures of risk aversion map into each other when the analysis controls for individual observed heterogeneity.

As it can be seen from Table 28, the interval regression estimations for all subjects in IP6 return a CRRA coefficient which is not significantly different from risk neutrality in the HL measure with high monetary stakes, significantly risk seeking in the HL measure with low monetary stakes, but significantly risk averse in the B-EG measure. The result for the HL task with high monetary payoffs is largely due to the high number of subjects who in the HL task with low monetary stakes choose lottery B in the first question, many of which then "switch back" to lottery A in the subsequent questions. When looking at the consistent subjects only, in fact, the interval regression estimations return a CRRA coefficient which is not significantly different from risk neutrality in both HL measures, while significantly risk averse in the B-EG measure. The same pattern emerges for the interval regressions for IP7 (available on request). These results suggest that the two experimental measures of risk aversion do not perfectly map into each other, a result that is in line with similar analysis by Loomes and Pogrebna (2014) and Crosetto and Filippin (2015). Also notice that our findings are in contrast with Dave et al. (2010) who, in a general population sample in Canada, find significantly higher risk aversion in the HL than in the B-EG task. In a companion paper (Galizzi et al., 2016a), we formally test this specific cross-validity aspect using maximum-likelihood structural estimations.

When a range of observable characteristics are introduced as controls in the interval regression estimations for the consistent subjects (Table 29), the estimated CRRA coefficient is not significantly different from risk neutrality in both HL measures, while it suggests significant risk aversion according to the B-EG measure. Female respondents and subjects with lower household per capita income are significantly more risk averse according to the B-EG measure, but not to the HL measures, while self-employed respondents are more risk taking according to the HL measure with high monetary stakes.

5.4.5 Ordered probit regression analysis: do the B-EG and HL measures of risk aversion predict the SOEP measures?

We next run a set of ordered probit regressions where the dependent variables are the SOEP risk attitudes measures: SOEP-G RA, for example, is defined as a decreasing measure of risk taking according to the SOEP-G task, taking value 1 (very willing to take risk) if the subject chooses 10 in the SOEP-G Likert scale ("I am generally a person fully prepared to take risks"), to value 10 (unwilling to take risk) if the subject chooses 0 in the SOEP-G Likert scale ("I am generally a person unwilling to take risks"). SOEP-F RA and SOEP-H RA are equivalently defined with the corresponding meaning for risk taking in finance and in health, respectively. In the set of ordered probit regressions the main explanatory variables are the CRRA values for the B-EG measures of risk aversion estimated by the interval regressions presented above, or the CRRA values for the HL measures of risk aversion with high and low monetary stakes estimated by the interval regressions presented above. Table 30 presents the ordered probit estimates for all subjects and for the consistent subjects only. Tables 31-34 include a set of further controls for individual characteristics. All ordered probit estimations allow for the complex survey design and use appropriate sampling weights to draw valid inference for the adult population in the UK, and adjust standard errors at strata and PSU levels.

From the ordered probit estimations in Table 30, it can be seen that, alike in the correlation analysis above, the B-EG measure of risk aversion is significantly and positively associated with the SOEP risk attitudes measure: the CRRA values estimated from the B-EG task associate with lower willingness to take risks in general, in finance, and in health. The associations are statistically significant and stronger when the analysis focuses on the consistent subjects who never switch back in the HL task.

The estimations for all respondents also show that, alike in the correlation analysis, the CRRA measure for the HL task with low monetary stakes is not significantly associated to any of the SOEP risk attitudes measures. Alike in the correlation analysis, the CRRA measure for the HL task with high monetary stakes is significantly associated with the SOEP-G measure. When, however, the analysis focuses on the consistent subjects only, there are significant positive associations between the SOEP-G and the CRRA measures for the HL tasks with both low and high monetary stakes.

For the consistent subjects, the CRRA measures for the HL tasks with low and high monetary stakes also associate marginally significantly with the SOEP-F, while there is no significant association between the CRRA measures for the HL tasks and the SOEP-H.

The next set of ordered probit regressions include as explanatory variables both the CRRA measures of risk aversion for the B-EG task and the HL task with low monetary stakes (Tables 31-32) or for the B-EG task and the HL task with high monetary stakes (Tables 33-34). This allows to test head-to-head the extent to which the B-EG or the HL measures of risk aversion are able to predict the SOEP measures. The estimations also add a further set of control variables in the ordered probit regressions. The regressions are conducted for all subjects and for the consistent subjects only.

The estimations for all respondents show that the CRRA measure for the B-EG task does a better job than the CRRA measure for the HL task with low monetary stakes in predicting the SOEP measures of risk taking, especially in finance and in health (Table 31): the CRRA values estimated from the B-EG task associate with lower levels of willingness to take risks in finance, in health, and marginally also in general. The association of the B-EG measure of risk aversion with the SOEP-G, however, is not robust to the introduction of further control variables. As discussed above, there is no significant association of the CRRA measure for the HL task with low monetary stakes with the SOEP-G when all respondents are considered in the analysis.

The estimations for all respondents also show that the CRRA measure for the HL task with high monetary stakes does a better job than the CRRA measure for the B-EG task in predicting the SOEP measures of risk taking in general, and the association is robust and stronger when the analysis controls for further explanatory variables (Table 31). When the B-EG and the HL CRRA measures are considered together, there is no significant association of the CRRA measure for the B-EG task with the SOEP-G when all respondents are considered in the analysis. The CRRA measure for the B-EG task, however, does a better job than the CRRA measure for the HL task with high monetary stakes in predicting the SOEP measures of risk taking in finance and to some extent also in health, and the association with the SOEP-F is robust to the introduction of further controls in the analysis.

Moreover, the estimations for the consistent subjects only show that the CRRA measures for the HL task with both high and low monetary stakes do a better job than the CRRA measure for the B-EG task in predicting the SOEP measures of risk taking in general, and that and the associations are robust and stronger when the analysis controls for further explanatory variables (Table 31 and 33). Also for the consistent subjects, the CRRA measure for the B-EG task does a better job than the CRRA measures for the HL tasks in predicting the SOEP measures of risk taking in finance and in health, and the associations with the SOEP-F and SOEP-H are robust to the introduction of further controls in the analysis.

As for the estimations that include further explanatory variables, three results are to be noted. First, as mentioned, the introduction of further explanatory variables does not alter the key cross-measures associations described above. In particular, even controlling for a broad range of variables and characteristics, the SOEP-G measure is significantly associated with the CRRA measure for the HL task with high monetary stakes and, for the consistent subjects, also with the CRRA measure for the HL task with low monetary stakes. On the other hand, the SOEP-F and SOEP-H measures are significantly associated with the CRRA measure for the B-EG task.

Second, age and gender are systematically associated to the SOEP risk attitudes measures. In all the estimations, older respondents tend to self-report significantly lower willingness to take risks in general, in finance, and in health, while female respondents tend to self-report significantly lower willingness to take risks in general and in finance. This set of results is generally in line with what documented for the SOEP measures by Josef et al. (2016) in a representative sample of the German population.

Third, for the consistent subjects, there is also an effect of the per-capita household income on the SOEP measures: consistent respondents with higher levels of per capita income report higher willingness to take risks in general, in finance, and in health. With the exception of the SOEP-H, however, the associations are not significant when all the respondents are considered. This set of results is also in line with what documented for a representative sample of the German population by Josef et al. (2016) who also do not find any significant association of income with responses to the SOEP questions.

The finding of a significant association between HL and SOEP risk attitudes measures for the consistent subjects is generally in line with the findings of Dohmen et al. (2011) and Falk et al. (2016) who use an incentive-compatible MPL test that imposes consistency of responses and find that it associates significantly with the responses to the SOEP-G task in their representative samples of the German population.

The lack of significant association between HL and SOEP-H measures of risk taking is broadly in line with the evidence that risk-taking could be largely domain-specific: see, for example, the discussion in Weber et al. (2002), Blais and Weber (2006), and the recent review of the evidence in Galizzi et al. (2016c).

5.4.6 Interval regression analysis: do the B-EG, HL, and SOEP measures predict each other?

We next systematically replicate the above cross-validity analysis using interval regressions. In particular, in Table 37 we run a set of cross-validity interval regressions where the dependent variable is the CRRA value conditional to the lottery choice in the B-EG task - estimated by the interval regressions illustrated in Section 5.4.4 - as a function of the SOEP risk attitudes measure, and of the CRRA values conditional to the specific switching points in the HL tasks - also estimated by the above interval regressions - considered either separately or together, and with and without the usual list of control variables. We similarly run corresponding interval regressions cross-validity estimations for the HL tasks, with low and high monetary stakes (Tables 35 and 36 respectively), where the dependent variable is the CRRA value conditional to the switching point in the HL task - estimated by interval regressions - as a function of the SOEP risk attitudes measure, and of the CRRA values conditional to the specific lottery choice in the B-EG task - also estimated by interval regressions. Our systematic approach allows us to complement the ordered probit estimations illustrated above to assess the robustness of the cross-validity associations across the different risk preferences measures.

As it can be seen in Table 37, there are positive significant associations between the

CRRA measure for the B-EG task and the CRRA measures for the HL tasks with both low and high monetary stakes. The associations are robust to the introduction of further control variables, and are stronger for the consistent respondents. There are positive significant associations also between the CRRA measure for the B-EG task and the SOEP risk attitude measures. The associations are stronger for the consistent respondents. The association of the SOEP-G with the CRRA measure for the B-EG task, however, is not robust to the introduction of further control variables in the interval regressions. As it can be seen in Tables 35 and 36, and coherently with the findings in Table 37, there are positive significant associations between the CRRA measures for the HL tasks with both low and high monetary stakes and the CRRA measure for the B-EG task. The associations are robust to the introduction of further control variables and are stronger for the consistent respondents. There is also a positive significant association of the SOEP-G with the CRRA measure for the HL task with high monetary stakes, which is robust to the introduction of further controls. The association of the SOEP-G with the CRRA measure for the HL task with low monetary stakes is only significant for the consistent subjects.

All in all, therefore, the results of the cross-validity interval regressions are in line with what found in the ordered probit regressions and in the pairwise correlations analysis.

We have also systematically replicated the analysis using a comprehensive set of alternative models and specifications. In particular, it is possible to associate each interval of CRRA values associated to a switching point in HL - or a lottery choice in B-EG - to the midpoint of that CRRA interval. The midpoints can be easily calculated for the closed-end CRRA ranges corresponding to the switching points 2 through 9 in the HL tasks and to the lottery choices B to E in the B-EG task.

Unlike for the switching points 2 through 9, the first and the last intervals of the CRRA ranges in the HL tasks (e.g. the one associated to choosing for the first time the riskier lottery B when $p = 0.1$ or to never choosing the riskier lottery B in the HL task) are open-ended. The same holds for the lottery choices A and F in the B-EG task. For the HL tasks, we have estimated the midpoints of the ranges of the two open-ended intervals in the HL tasks as the fitted values of the regression of the midpoints as a function of the switching points:

$$\text{Midpoint}_i = \beta_0 + \beta_1 \text{Switching P}_i + \beta_2 \text{Switching P}_i^2 + \beta_3 \text{Switching P}_i^3 + u_i$$

The function is an approximation of an out-of-sample extension of the locally weighted regression of the midpoints on the switching points (i.e. the "lowess" smoother), as show in Figure 2.

Similarly, for the B-EG task we have estimated the midpoints of the ranges of the two open-ended intervals as the fitted values of the regression of the midpoints as a function of the lottery choices:

$$\text{Midpoint}_i = \beta_0 + \beta_1 \text{Lottery}_i + \beta_2 \text{Lottery}_i^2 + \beta_3 \text{Lottery}_i^3 + \nu_i$$

Also this function is an approximation of an out-of-sample extension of the locally weighted "smoother" regression of the midpoints on the lottery choices, as show in Figure

3.[14]

We have then run a set of regressions to replicate the cross-validity regressions using the corresponding midpoints of the range of the CRRA values for the B-EG and the HL tasks. In particular, in a set of ordered probit regressions, the dependent variable are the SOEP risk attitudes measures while the independent variables are the calculated midpoints of the ranges of the CRRA values associated to the lottery choices in the B-EG and HL tasks, considered either separately or together. In another set of linear regressions, the dependent variables are the calculated midpoints of the ranges of the CRRA values associated to the responses to the B-EG task, while the independent variables are the SOEP risk attitude measure and the midpoints of the ranges of the CRRA values in the HL task, considered either separately or together. A third set of linear regressions has been equivalently conducted with the dependent variables being the midpoints of the ranges of CRRA values for the HL task. We have also run interval regressions for the CRRA value conditional to the lottery choice in the B-EG task - estimated by the above interval regressions - as a function of the SOEP risk attitudes measures and of the midpoint of the CRRA ranges for the HL task, again considered either separately or together, and similarly for the interval regressions where the dependent variables are the CRRA measures for the HL tasks. All sets of regressions have been replicated with and without the usual list of control variables, and considering either the HL task with low or high monetary stakes. All estimations allow for the complex survey design, and for the fact that sampling weights for each subject reflect the adult population in the UK, and adjust standard errors at strata and PSU level.

We have found the same patterns of significant associations across the interval regressions presented above and this further set of regressions using the midpoints of the CRRA ranges. The results of these further cross-validity regressions are all available on request.

5.5 Temporal stability

We next look at within-subject responses at two points in time and describe the behavior of the subsample of subjects who answered the same set of risk preferences questions in both IP6 and IP7, a year after.

5.5.1 Temporal stability of B-EG responses

A total of 406 subjects responded to the B-EG questions in both IP6 and IP7. For these subjects, the whole set of choices in IP6 and IP7 is reported in Table 11 where we describe the entire distribution of the ranges of CRRA values associated to each lottery choice in the B-EG task in IP6. If individual choices are fully stable across the two waves, one should expect the proportions along the main diagonal to be 100%.

[14]We have compared the calculated midpoints of the different ranges of the CRRA with the CRRA values estimated by interval regressions for the B-EG and HL tasks. As it can be seen from Figures 2-3, in both the B-EG and the HL task, the midpoints of the ranges overlap with the CRRA values estimated by interval regressions for the closed intervals. In the left open-ended ranges of the B-EG and HL tasks, the calculated midpoints overestimate the CRRA values compared to the interval regressions, while they underestimate the CRRA values in the right open-ended ranges.

Out of the 409 subjects who responded to the B-EG questions in both IP6 and IP7, only 29.55% chose the same B-EG lottery option in both waves. This proportion is larger for the subjects who in IP6 preferred lottery A, the safe lottery (46.3%). The proportion of temporally stable choices is slightly higher (32.46%) for the sub-sample of 285 respondents who always made consistent choices in the HL questions in IP6, thus excluding the subjects who switched back. The overall set of B-EG choices in IP6 and IP7 for all the subjects and for the subset of subjects consistent in the HL IP6 task is also described in Table 11.

5.5.2 Temporal stability of HL responses

In principle, if individual choices are fully stable across the two waves, one should expect to see, for each individual, exactly the same series of 18 binary choices in both IP6 and IP7. This seems realistically a too strong definition of stability across waves. An alternative is to look at the first point in the sequence of choices where each subject switches from lottery A to lottery B in the block of 9 HL questions with low monetary stakes, and in the block of 9 HL questions with high monetary stakes. In the case of the consistent subjects, this switching point is unique and directly identifies the underlying range of CRRA values, increasing with risk aversion, giving a precise measure of cross-waves stability.[15]

A total of 407 subjects responded to the HL lotteries with low monetary stakes in both IP6 and IP7. For these subjects, Table 12 reports the whole distribution of the ranges of CRRA values associated with each first switching point in the HL task with low monetary stakes in IP6 and IP7. If individual choices are fully stable across the two waves, one should expect again the proportions along the main diagonal to be 100%.

Out of the 407 subjects who responded to the HL questions with low monetary stakes in both IP6 and IP7, only 29.49% switched for the first time to lottery B in correspondence of the same point in both waves. The proportion of temporally stable responses is larger for the subjects who in IP6 chose lottery B immediately (52.96%), or who never switched to lottery B (37.87%).[16]

For the HL choices with low monetary stakes there are several cases of local movements (e.g. 43.56% of the subjects who chose to first switch to lottery B in the second binary choice in IP6 switched immediately to lottery B in IP7). The proportion of temporally stable responses is similar (26.83%) for the 282 consistent subjects (Table 12, Panel B).

Analogous results hold for the HL questions with high monetary stakes (Table 13). Out of the 406 subjects who responded to the HL questions with high monetary stakes in both IP6 and IP7, only 30.73% switched for the first time to lottery B in correspondence of the same point in both waves.

[15]In the case of inconsistent subjects who switch back, clearly this switching point is just the first of multiple switching points, but it could still be used as a tentative metrics for the stability of responses across waves.

[16]Just as a comparison, Hey and Orme (1994) find an average consistency rate of around 75% across two MPL sessions at a distance of a week to 10 days from each other.

5.5.3 Temporal stability of SOEP responses

A total of 355 subjects responded to the SOEP-G and SOEP-F questions in both IP6 and IP7, while 354 subjects also responded to the SOEP-H questions in both waves. For these subjects, the whole set of choices in IP6 and IP7 is reported in Tables 14-16 where we describe the entire distribution of the SOEP risk attitudes measures in IP7 for each value of the SOEP risk attitude measures in IP6.

Out of the 355 subjects who responded to the SOEP-G questions in both IP6 and IP7, only 24.24% rated their willingness to take risk in general in exactly the same way in both waves. These proportions are higher when in IP6 the subjects self-reported no willingness to take risks (28.18%), maximum willingness to take risks (46.5%) or a willingness to take risks of 5, the midpoint of the scale (40.22%). There is, moreover, a visible pattern of choices around the main diagonal of the table, with evidence of local movements in responses from IP6 to IP7. There are some visible patterns of choices at or around the main diagonal of the tables also in the responses to the SOEP-F and SOEP-H questions.

5.5.4 Interval regression and ordered probit analysis: do the IP6 measures predict the IP7 measures?

We next systematically replicate the above cross-waves correlational analysis using interval regressions and ordered probit regressions. In particular, in Table 38 we run a temporal stability interval regressions where the dependent variables are the CRRA intervals associated to the lottery choice in the B-EG task in IP7 as functions of the CRRA value conditional to the lottery choice in the B-EG task in IP6 estimated by the interval regressions illustrated in Section 5.4.4.

We similarly run corresponding interval regressions estimations for the temporal stability of the HL measures, with low and high monetary stakes (Table 39), where the dependent variables are the CRRA intervals associated to the switching points in the HL tasks in IP7 as a functions of the CRRA values conditional to the switching points in the HL tasks in IP6.

We also run a set of ordered probit regressions where the dependent variables are the SOEP risk attitude measures in IP7 - defined as decreasing measures of risk taking attitudes in general, in finance, and in health according to the SOEP-G, SOEP-F, and SOEP-H tasks, respectively, as functions of the corresponding SOEP measures in IP6 (Table 40).

All ordered probit estimations and interval regressions are replicated with the usual list of control variables, for all subjects and for the consistent subjects only. This set of ordered probit and interval regressions allows to replicate the temporal stability correlational analysis in a regression framework and to systematically look at whether the different measures of risk preferences in IP6 can predict the corresponding risk preferences measures in IP7 when the analysis controls for individual observed heterogeneity.

From the interval regressions estimations in Table 38, it can be seen that, alike in the correlation analysis above, the B-EG measure of risk aversion in IP6 is significantly and positively associated with the B-EG measures of risk aversion in IP7 and that the effect is robust to the introduction of further control variables and when focusing on the consistent subjects only. Similarly, the interval regressions estimations in Table 39 show that, alike in the correlation analysis, the HL measure of risk aversion in IP6 is significantly and positively associated with the HL measures of risk aversion in IP7 and that the effect is robust to the introduction of further control variables and when focusing on the consistent subjects only. Finally, the ordered probit estimations in Table 40 suggest that SOEP risk attitudes measures in IP6 are significantly associated with the corresponding SOEP measures in IP7, and the effects are robust to the introduction of further control variables and when focusing on the consistent subjects only.

5.6 External validity

We then look at the issue of external validity of the three measures of risk preferences. To do so, we conduct a systematic external validity analysis to predict a range of health and financial behaviours using ordered probit or linear regression models, as appropriate. The dependent variables are a set of individual behaviours from the linked UKHLS data which may be conceptually associated to individual risk preferences. We focus, in particular, on health and financial behaviours. For health behaviours, we look at the questions about smoking or not smoking, and about heavy drinking, repeated in both IP6 and IP.

We also conduct wave-specific sets of regressions taking advantage of the fact that in IP6 we have variables for the BMI, and variables for eating fast food, and for whether the respondents are regular savers, have a private pension fund, while in IP7 we have information on consumption of fruit and vegetables, and on the DOSPERT test for domain-specific hypothetical risk taking situations.

The main explanatory variables are the SOEP risk attitude measures, and the CRRA values conditional to the specific lottery choice in the B-EG task and to the specific switching point in the HL task estimated by the above interval regressions.

All the external validity regressions have been replicated using a comprehensive set of alternative models and specifications using the midpoints of the CRRA ranges instead of the CRRA values estimated by the interval regressions.[17] We have found the same patterns of significant associations across the further sets of regressions using the midpoints of the CRRA ranges and the ones using the CRRA values estimated by the interval regressions. In what follows, therefore, we present the results of the latter estimations, while the results of the former are all available on request.

As it can be seen from Table 41, none of the B-EG and HL measures of risk aversion and SOEP risk attitude measures are associated to the respondents' smoking status in IP6. In IP7 there is some associations of the SOEP-G and SOEP-H measures with the smoking status but they become only marginally significant when the analysis controls for other individual characteristics.

As it can be seen in Table 42, there is a significant and positive association between the BMI and the CRRA measures for the HL tasks with high monetary stakes and for

[17]All external validity estimations allow for the complex survey design, and use appropriate sampling weights to draw valid inference for the adult population in the UK, and adjust standard errors at strata and PSU levels.

the B-EG task in IP6: subjects with higher BMI tend to be more risk averse. These associations are robust to the introduction of control variables in the analysis. There is no significant association between the BMI and any of the SOEP measures.

As it can be seen in Table 43, none of the B-EG and HL measures of risk aversion and SOEP risk attitude measures are associated to the respondents' consumption of fast food in IP6. The marginally significant effect of the SOEP-G is not robust to the introduction of control variables in the analysis.

On the other hand, as seen in Table 44, the CRRA measures for the HL task with high monetary stakes and (marginally) with low monetary stakes are positively and significantly associated with the regular consumption of fruit and vegetables in IP7, and these associations are robust to the introduction of control variables in the analysis. The associations of the SOEP-F and SOEP-H with fruit and vegetable consumption is not robust to the introduction of control variables.

At the contrary, the SOEP-G and SOEP-H risk attitudes measures are negatively and significantly associated with a variable indicating how many times the respondents had five or more alcoholic drinks in the last four weeks, although the associations are not robust to the introduction of control variables in the analysis. The CRRA measures for the B-EG and HL tasks are not associated with the drinking status.

Moving to financial behaviour, as it can be seen in Tables 48 and 49, none of the B-EG and HL measures of risk aversion, and SOEP risk attitude measures are associated to whether the respondents regularly save, nor to the composition of the savings between long term and short term time horizons.

On the other hand, the SOEP-G risk attitude measure is positively and significantly associated with the log of the respondents' savings, and the effect is robust to the introduction of control variables in the analysis. Furthermore, as it can be seen in Table 51, the CRRA measure for the B-EG task is negatively significantly associated with the respondents' likelihood of having a private pension fund, and the association is robust to the introduction of controls in the analysis: the subjects who, according to the B-EG task, are more risk seeking in their lottery choices, are more likely to have a private pension fund.

Finally, as seen in Tables 52 and 53, all the three SOEP risk attitudes measures are negatively and significantly associated with the summary scores of hypothetical risk-taking behaviour in finance and in health as measured by the DOSPERT test, and the associations are robust to the introduction of control variables in the analysis. There is no significant association between the DOSPERT scores of risk taking in finance and in health and the CRRA measures for the B-EG and HL tasks.

The main findings above are replicated when the external validity regressions are conducted focusing on the consistent subjects only (all results available at request). In addition, there are other significant associations: the BMI is positively significantly associated with the CRRA measures for both high and low monetary stakes; the smoking status is marginally negatively significantly associated with the SOEP-G measure; the DOSPERT score for hypothetical risk-taking in finance is negatively significantly associated with the SOEP-G and SOEP-F measures, but not with the SOEP-H measure.

The positive association between the HL measure of risk aversion and the BMI is in contrast with Anderson and Mellor (2008) who, in a sample of US adults, found a negative relationship between the HL switching point and a BMI indicator for being overweight: subjects with higher BMI were *less* risk averse, although the effect was not robust to changes in the BMI threshold used to define overweight or obese subjects. The finding that risk preferences are significantly associated with nutritional behavior, such as the consumption of fruit and vegetables, is in line with similar findings by Galizzi and Miraldo (2012) calculating the Healthy Eating Index (HEI) for a sample of UK students. Moreover, the lack of association between smoking and risk preferences measures adds to the evidence cumulating from a number of studies that have failed to find significant correlations between risk preferences and smoking (Yi et al., 2007; Reynolds et al., 2007; Harrison et al., 2010; Mitchell, 1999; Galizzi and Miraldo, 2012; Szrek et al., 2012; Harrison et al., 2015a).

6 Inconsistent subjects: a closer look

In the previous sections we have reported the results on the cross-validity and external validity of the different B-EG and HL measures of risk aversion and SOEP risk attitude measures for the whole sample of respondents and focusing only on the group of "consistent" subjects who never behave inconsistently in the HL questions.

As described in Section 5.2.1, there is a total of 206 subjects who behaved inconsistently at least once in the HL questions in IP6, in the sense that they switched back to lottery A after choosing lottery B at least once in the HL task.

In the above analysis we have pooled all these inconsistent subjects together in the same group, and we have conservatively excluded them when looking at the choices of the consistent subjects only. One may wonder, however, whether some more information on the cross-validity, temporal stability, and external validity of the three measures of risk aversion can be extracted from at least some of these inconsistent subjects. After all, the degree to which the behaviour can be considered inconsistent can vary across different subjects. In this section we give a closer look at these inconsistent subjects and we try to unpack them in separate "types" according to how much we can plausibly learn from their behaviour.[18]

A total of 160 and 126 subjects made at least one inconsistent choice when answering HL questions with low or high monetary stakes, respectively, in the sense of switching back to lottery A after having chosen lottery B. Moreover, a total of 80 subjects made at least one inconsistent choice in both the low and the high monetary stakes HL questions.

Among the 160 subjects who made at least one inconsistent choice when answering HL questions with low monetary stakes, 113 subjects switched from lottery B back to lottery A exactly one time, 39 switched back two times, 6 switched back three times, and 2 have missing values.

Among the 126 subjects who made at least one inconsistent choice when answering HL questions with high monetary stakes, 85 subjects switched from lottery B back to lottery

[18]For another study looking in greater detail at the behaviour of inconsistent subjects in the HL task, see Jacobson and Petrie (2009).

A exactly one time, 24 switched back two times, 11 switched back three times, 3 switched back four times, and 3 have missing values.

In what follows, we focus on the inconsistent subjects who in the HL questions with low monetary stakes (the ones presented first) switched from lottery B back to lottery A only one time. These subjects can arguably be considered the "least inconsistent" among all the inconsistent subjects. In particular, three groups of such subjects can be identified.

The first group are all subjects who chose lottery B in the very first HL question with low monetary stakes but then chose lottery A in all other subsequent HL questions with low monetary stakes. We call these subjects "learners" as they could have made a "mistake" in their choices in the first question, but then learned and otherwise made consistent choices. In particular, there are 14 subjects who chose lottery B in the very first HL question with low monetary stakes, but then consistently chose lottery A in all other subsequent HL questions with low monetary stakes. That these 14 subjects are indeed learning is confirmed by the fact 9 out of these 14 subjects then behave fully consistently in their choices in the HL questions with high monetary stakes.

The second group are all subjects who always chose lottery A in the HL questions except at one single question where they chose lottery B but then chose lottery A again in all subsequent choices. We call these subjects as "trembling hand risk averse" as they could have made a mistake at one question, but otherwise made consistently risk averse choices. There are 9 of these trembling hand risk averse subjects in the HL questions with low monetary stakes. That also these subjects are indeed learning is confirmed by the fact that 5 of these 9 subjects then behave fully consistently in their choices in the HL questions with high monetary stakes.

The last group are all subjects who always chose lottery B in the HL questions except at one single question where they chose lottery A but then chose lottery B again in all subsequent choices. We call these subjects as "trembling hand risk seeking" as they could have made a mistake at one question, but otherwise made consistently risk seeking choices. There are 9 of these trembling hand risk seeking subjects in the HL questions with low monetary stakes. That also these subjects are indeed learning is confirmed by the fact that as many as 7 of these 9 subjects then behave fully consistently in their choices in the HL questions with high monetary stakes.

In what follows, we consider the subjects in the above three groups and we tentatively treat them as only "locally inconsistent": these subjects made only one single "mistake" in the sequence of HL choices with low monetary stakes, but their choices are otherwise consistent.[19]

In this section, therefore, we temporarily disregard the "mistake" that these "locally inconsistent" subjects could have made in their HL choices, and we tentatively estimate, using interval regressions, a CRRA measure for their HL task with low monetary stakes in the same way as we illustrate above for the consistent subjects only.

[19] One can of course define further potential groups or "types" of inconsistent subjects who could be plausibly considered as making at least some consistent choices. We have looked into each individual series of HL choices searching for more learning patterns, but the definition and categorization of further "types" risks to become more blurred, ad-hoc, and somewhat arbitrary. We have therefore conservatively opted to consider all remaining subjects as "truly inconsistent" and we have pooled all of them together.

In particular, for each of the "learners" subjects, we disregard the "mistake" in the very first HL questions and we estimate, using interval regressions, a CRRA value based on the hypothesis that these subjects are otherwise consistently risk averse.

Similarly, for each of the "trembling hand risk averse" and "trembling hand risk seeking" subjects, we disregard the "trembling hand" mistake and we estimate, using interval regressions, a CRRA value based on the hypothesis that they were otherwise consistently risk averse and risk seeking, respectively.

We then pool all these "locally inconsistent" subjects together with the "consistent" ones and we replicate the analysis that we have described above for the consistent subjects only, in order to see the extent to which results are similar to the ones illustrated for the sample of the consistent subjects only.

Overall, a total of 37 subjects could be considered "locally inconsistent" according to the criteria explained above. There are therefore up to 457 subjects in the pooled sample of consistent and "locally inconsistent" subjects. We next present the cross-validity and external validity analyses for this pooled sample and we compare the findings with the corresponding ones presented above for the whole sample and for the subsample of the consistent respondents.

As it can be seen from Table 54, when the "locally consistent" are pooled together with the consistent subjects, the ordered probit cross-validity estimations return substantially the same results as the ones described above. In particular, the CRRA measure for the HL task with high monetary stakes, but not with low monetary stakes, is significantly and positively associated with the SOEP-G measure. Moreover, the B-EG measure of risk aversion is significantly and positively associated with all the three SOEP measures (Table 55).

Also the interval regressions cross-validity estimations with the "locally consistent" pooled together with the consistent subjects return substantially the same results as the ones described above. In particular, there are positive significant associations between the CRRA measure for the B-EG task and the CRRA measures for the HL tasks with both and high monetary stakes, and the associations are robust to the introduction of controls in the analysis (Tables 56-57). There are positive significant associations also between the CRRA measure for the B-EG task and the SOEP risk attitudes measures. The association of the SOEP-G with the CRRA measure for the B-EG task, however, is not robust to the introduction of further control variables in the interval regressions.

Furthermore, the interval regressions cross-validity estimations with the "locally consistent" pooled together with the consistent subjects confirm that, coherently with the findings above, there are positive significant associations between the CRRA measures for the HL tasks with both and high monetary stakes and the CRRA measure for the B-EG task, and that the associations are robust to the introduction of further control variables (Tables 58-59). Moreover, the SOEP-G and the SOEP-F measures are significantly and positively associated with the CRRA measure for the HL task with high monetary stakes, and the associations are robust to the introduction of controls in the analysis. The SOEP-G measure is only marginally significantly associated with the CRRA measure for the HL task with low monetary stakes. Analogous temporal stability and external validity estimations provide substantially similar results as the ones discussed above for the sample of consistent subjects only (tables available on request).

All in all, therefore, the results of the cross-validity, temporal stability, and external validity regressions for the pooled sample of consistent and "locally inconsistent" subjects are in line with the above findings for the consistent subjects only. We have also systematically replicated the analysis using a comprehensive set of alternative models and specifications using either the raw responses or the midpoints of the CRRA ranges and found again substantially identical results (all available on request).

7 Conclusions

Risk preferences are considered a fundamental driver of individual behavior in a broad array of contexts, including health, social care, education, migration, occupational and self-employment choices, personal and household finance (Barsky et al., 1997; Gollier, 2001; Bonin et al., 2007; Harrison and Rutström, 2008; Anderson and Mellor, 2008; Guiso and Paiella, 2008; Bucciol and Miniaci, 2011; Von Gaudecker et al., 2011; Bellemare and Shearer, 2013; Charness et al., 2013; Sutter et al., 2013). Experimental and behavioral economists have developed a range of different methods to measure risk preferences, including the HL (Holt and Laury, 2002), the B-EG (Binswanger, 1980, 1981; Eckel and Grossman, 2008), and the SOEP method (Dohmen et al., 2011).

Using a UK nationally representative sample of adult respondents within the Innovation Panel of the UKHLS, we have systematically looked at the stability and validity of the B-EG and HL experimental measures of risk preferences, and of the SOEP risk attitudes questions along three directions: at a given point in time across different methods (*cross-validity*); over time (*temporal stability*); and in terms of the degree to which they associate with a range of field behaviors (*external validity*).

To the best of our knowledge, our study is the first to look jointly and systematically at the three validity dimensions of two experimental measures of risk preferences (B-EG and HL) and three risk attitudes questions (SOEP-G, SOEP-F, and SOEP-H) using a nationally representative sample. Looking systematically and comprehensively at all the three validity dimensions together is important because it allows us to draw conclusions on the *overall validity* of different risk preferences measures that would otherwise be missed when looking at each aspect in isolation.

We have three main findings. First, concerning *cross-validity*, we find evidence that the different risk preferences measures generally correlate and map into each other, although their associations are not perfect. In particular, the two experimental measures are correlated to each other; the HL experimental measure with high monetary stakes is significantly associated with the survey question about self-reported risk attitude in general but not with the self-reported risk attitudes in the financial and health domains; the HL experimental measure with low monetary stakes is not correlated with any of the questions on self-reported risk attitudes; the B-EG experimental measure is instead significantly associated to all of the survey questions about risk attitudes.

Second, concerning *temporal stability*, we find that, after one year, less than one third of the subjects choose the same preferred option in the lottery choices or in the self-reported risk attitude scales. Nonetheless, there is a significant and positive correlation between the risk aversion parameters implicit in the lottery choices in IP6 and IP7, as well as between the self-reported risk attitudes in the two waves. Overall, the temporal stability of the risk attitudes question is higher than that of the experimental measures. Moreover, the parameters of risk aversion implicit in the B-EG and HL lottery choices in IP6 and the self-reported attitudes in IP6 significantly predict the corresponding risk aversion parameters and self-reported attitudes in IP7.

Third, we find mixed evidence concerning the *external validity* of the three risk aversion measures. In particular, none of the three measures of risk aversion is associated with the respondents' smoking status, their consumption of junk food, their likelihood of regularly saving, or the time horizons of their savings. The two experimental measures significantly associate with the subjects' BMI, whereas none of the survey questions do. Only the HL and B-EG measure predict the regular consumption of fruit and vegetables and whether the respondents have a private pension fund, respectively; on the other hand, only the survey questions associate with the DOSPERT scores for risk-taking in hypothetical situations in finance and in health, and with heavy alcohol drinking.

All in all, the evidence on the *overall validity* of the three risk preferences measures is rather mixed. On the one hand, all three measures show strong and significant test-retest correlations across the two subsequent waves of data collection. Each of the measures, moreover, seem to significantly associate to some risky behavior in the field: the B-EG and HL measures associate with having a private pension, the HL with the BMI and the consumption of fruit and vegetables, while the SOEP measures associate with a range of hypothetical risk-taking situations in health and finance.

On the other hand, none of the three measures of risk aversion is significantly associated with variables of key health policy interest such as the respondents' smoking status or the consumption of junk food. More generally, with the exception of the BMI, no risky health behaviour is associated with the B-EG measure of risk aversion, and few of them are associated with the SOEP measures. The associations of the three risk aversion measures with financial behaviours are also not systematic. Moreover, the way in which the three different measures of risk aversion correlate and map into each other is far from being perfect.

From this perspective, our systematic study with a representative sample of the UK population adds to the mixed evidence on the cross-validity and external validity of different experimental measures of risk preferences (Isaac and James, 2000; Anderson and Mellor, 2009; Dave et al., 2010; Reynaud and Couture, 2012; Deck et al., 2013; Dulleck et al., 2013; Loomes and Pogrebna, 2014; Crosetto and Filippin, 2015; Vieider et al., 2015b; Attanasi et al., 2016). Caution should thus be in order when using one specific risk preferences measure to draw conclusions about the various, inevitably multifaceted, dimensions of individual risk-taking in different contexts, especially for representative samples of the population that, unlike conventional student subjects, are highly heterogeneous.

At the same time, our longitudinal artefactual field experiment with a representative sample of the UK population adds to the evidence on the temporal stability of risk preferences (Andersen et al., 2008a; Harrison and Swarthout, 2014; Josef et al., 2016).

Of course, further research is needed to test whether similar conclusions hold for more general models of risk preferences beyond the reference EUT and CRRA cases considered here (Hey and Orme, 1994; Bruhin et al., 2010; Wakker, 2010); when risk preferences are structurally estimated jointly with time preferences and using mixtures models (Andersen et al., 2008b, 2014; Harrison et al., 2009; Conte et al., 2011; Balcombe and Fraser, 2015; Galizzi et al., 2016b); or when the analysis account for the fact that respondents in a representative sample like ours integrate experimental prizes within individual-specific levels of background income (Andersen et al., 2015; Galizzi et al., 2016a).

Acknowledgement

We make use of data from the Understanding Society Innovation Panel administered by ISER, University of Essex, and funded by the Economic and Social Research Council. We also make use of risk and time preferences data collected as part of the Future Research Leader project Linking Experimental and Survey Data: Behavioural Experiments in Health and Wellbeing (ES/K001965/1, PI: MM Galizzi), funded by the Economic and Social Research Council (ESRC). We are grateful to the ESRC for financial support. We thank Nick Buck, Jon Burton, Annette Jackle, Jakob Petersen, and Noah Uhrig, and the ISER for help in implementing the experimental tests. We gratefully thank Stephen Jenkins for help and support in the proposal of linking experiments with Understanding Society survey data. We thank Paolo Crosetto and Antonio Filippin for sharing their information on the proportion of "inconsistent" respondents in previous experiments with the HL task. We gratefully thank Glenn Harrison and Peter Wakker for detailed feedback and discussion on an earlier version of this work. We thank Kelvin Balcombe, Aurelien Bailon, Nick Bardsley, Michaela Benzeval, Philip Corr, Robin Cubitt, Randy Ellis, Simon Gachter, John Gathergood, Nikos Georgantzis, Stephen Jenkins, Leonhard Lades, Morten Lau, Tiziana Leone, Albert Ma, Antoine Nebout, Jawwad Noor, Chiara Orsini, Aurora Ortiz-Nunez, Berkay Ozcan, Daniele Paserman, Lucida Platt, Johannes Schmieder, Amanda Sheely, Chris Starmer, Fabio Tufano, Joerg Weber, Eugenio Zucchelli, and all participants in the EuHEA conference at University of Hamburg, the FUR conference at University of Warwick, the NIBS conference at University of East Anglia, and in seminars at the universities of City London, Copenhagen University, Erasmus University Rotterdam, Imperial College, Lancaster, Nottingham, and Reading for helpful comments and suggestions.

References

Andersen, S., Cox, J. C., Harrison, G. W., Lau, M. I., Rutström, E. E., and Sadiraj, V. (2015). Asset integration and attitudes to risk: theory and evidence. Working paper, CEAR.

Andersen, S., Harrison, G. W., Lau, M. I., and Elisabet Rutström, E. (2008a). Lost in state space: Are preferences stable? *International Economic Review*, 49(3):1091–1112.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008b). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2014). Discounting behavior: A reconsideration. *European Economic Review*, 71:15–33.

Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, 27(5):1260–1274.

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

Attanasi, G., Georgantzís, N., Rotondi, V., Vigani, D., et al. (2016). Lottery-and survey-based risk attitudes linked through a multichoice elicitation task. Technical report, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg.

Balcombe, K. and Fraser, I. (2015). Parametric preference functionals under risk in the gain domain: A bayesian analysis. *Journal of Risk and Uncertainty*, 50(2):161–187.

Baltussen, G., Post, G. T., Van Den Assem, M. J., and Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. *Experimental Economics*, 15(3):418–443.

Barseghyan, L., Prince, J., and Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? evidence from insurance data. *American Economic Review*, 101(2):591–631.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2):537–579.

Baucells, M. and Villasís, A. (2010). Stability of risk preferences and the reflection effect of prospect theory. *Theory and Decision*, 68(1):193–211.

Beattie, J. and Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, 14(2):155–168.

Becker, A., Dohmen, T. J., Enke, B., Falk, A., Huffman, D., Sunde, U., et al. (2015). The nature and predictive power of preferences: Global evidence. Discussion paper, CEPR Discussion Papers.

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

Bellemare, C. and Shearer, B. (2013). Multidimensional heterogeneity and the economic importance of risk and matching: evidence from contractual data and field experiments. *The RAND Journal of Economics*, 44(2):361–389.

Berg, J., Dickhaut, J., and McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, 102(11):4209–4214.

Bernasconi, M. (1994). Nonlinear preferences and two-stage lotteries: Theories and evidence. *The Economic Journal*, pages 54–70.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics*, 62(3):395–407.

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural india. *The Economic Journal*, 91(364):867–890.

Blais, A.-R. and Weber, E. U. (2006). A domain-specific risk-taking (dospert) scale for adult populations. *Judgment and Decision Making*, 1(1).

Bonin, H., Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14(6):926–937.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412.

Bucciol, A. and Miniaci, R. (2011). Household portfolios and implicit risk preference. *Review of Economics and Statistics*, 93(4):1235–1250.

Chambers, R. G. and Quiggin, J. (2000). *Uncertainty, production, choice, and agency: the state-contingent approach*. Cambridge University Press.

Charness, G. and Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, 48(1):133–146.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51.

Charness, G. and Viceisza, A. (2015). Three risk-elicitation methods in the field: Evidence from rural senegal. *Review of Behavioral Economics*, page (in press).

Chuang, Y. and Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, 117:151–170.

Conlisk, J. (1989). Three variants on the allais example. *The American Economic Review*, pages 392–407.

Conte, A., Hey, J. D., and Moffatt, P. G. (2011). Mixture models of choice under risk. *Journal of Econometrics*, 162(1):79–88.

Cox, J. C. and Harrison, G. W., editors (2008). *Risk aversion in experiments*. Emerald Group Publishing, Bingley, UK.

Cox, J. C. and Sadiraj, V. (2006). Small-and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior*, 56(1):45–60.

Cox, J. C., Sadiraj, V., and Schmidt, U. (2015). Paradoxes and mechanisms for choice under risk. *Experimental Economics*, 18(2):215–250.

Cox, J. C., Smith, V. L., and Walker, J. M. (1988). Theory and individual behavior of first-price auctions. *Journal of Risk and uncertainty*, 1(1):61–99.

Crosetto, P. and Filippin, A. (2013). The bomb risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Crosetto, P. and Filippin, A. (2015). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, pages 1–29.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, 1(2):115–131.

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

Deck, C., Lee, J., Reyes, J. A., and Rosen, C. C. (2013). A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. *Journal of Economic Behavior & Organization*, 87:1–24.

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

Donkers, B., Melenberg, B., and Van Soest, A. (2001). Estimating risk attitudes using lotteries: A large sample approach. *Journal of Risk and Uncertainty*, 22(2):165–195.

Drehmann, M., Oechssler, J., and Roider, A. (2005). Herding and contrarian behavior in financial markets: An internet experiment. *The American economic review*, 95(5):1403–1426.

Dulleck, U., Fooken, J., and Fell, J. (2013). Within-subject intra-and inter-method consistency of two experimental risk attitude elicitation methods. *German Economic Review*, 16(1):104–121.

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.

Einav, L., Finkelstein, A., Pascu, I., and Cullen, M. R. (2012). How general are risk preferences? choices under uncertainty in different domains. *American Economic Review*, 102(6):2606–38.

Evans, W. N. and Viscusi, W. K. (1991). Estimation of state-dependent utility functions using survey data. *The Review of Economics and Statistics*, 73(1):94–104.

Falk, A., Becker, A., Dohmen, T. J., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Discussion paper, IZA Discussion Paper.

Galizzi, M. M., Harrison, G. W., Machado, S., and Miniaci, R. (2016a). The role of background income in estimating risk preferences: evidence from a uk representative sample. Discussion paper, Unpublished manuscript.

Galizzi, M. M., Harrison, G. W., and Miniaci, R. (2016b). Linking experimental and survey data for a uk representative sample: structural estimation of risk and time preferences. Discussion paper, Unpublished manuscript.

Galizzi, M. M. and Miraldo, M. (2012). Are you what you eat? experimental evidence on health habits and risk preferences. Discussion paper 2012-04, Imperial College Business School.

Galizzi, M. M., Miraldo, M., and Stavropoulou, C. (2016c). In sickness but not in wealth: field evidence on patients? risk preferences in the financial and health domain. *Medical Decision Making*, 36(4):503–517.

Galizzi, M. M. and Navarro-Martinez, D. (2015). On the external validity of social-preference games: A systematic lab-field study. Working Paper 802, Barcelona GSE.

Georgantzís, N. and Navarro-Martínez, D. (2010). Understanding the wta–wtp gap: Attitudes, feelings, uncertainty and personality. *Journal of Economic Psychology*, 31(6):895–907.

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

Gollier, C. (2001). *The Economics of Risk and Time*. The MIT Press.

Guiso, L. and Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic association*, 6(6):1109–1150.

Guiso, L., Sapienza, P., and Zingales, L. (2013). Time varying risk aversion. Discussion paper, National Bureau of Economic Research.

Hanoch, Y., Johnson, J. G., and Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment an application to risk-taking behavior. *Psychological Science*, 17(4):300–304.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, 5(1):53–84.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2010). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 120(545):595–611.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, pages 1251–1289.

Harrison, G. W. (1986). An experimental test for risk aversion. *Economics Letters*, 21(1):7–11.

Harrison, G. W. (1990). Risk attitudes in first-price auction experiments: A bayesian analysis. *The Review of Economics and Statistics*, pages 541–546.

Harrison, G. W., Hofmeyr, A., Ross, D., and Swarthout, J. T. (2015a). Risk preferences, time preferences and smoking behaviour. Paper presented at the Third Workshop on Behavioural and Experimental Health Economics.

Harrison, G. W., Johnson, E., McInnes, M. M., and Rutström, E. E. (2005a). Risk aversion and incentive effects: Comment. *American Economic Review*, pages 897–901.

Harrison, G. W., Johnson, E., McInnes, M. M., and Rutström, E. E. (2005b). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1(1):31–35.

Harrison, G. W. and Lau, M. I. (2014). Risk attitudes, sample selection, and attrition in a longitudinal field experiment. CEAR WP 2014-04, CEAR, Georgia State University.

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

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization*, 70(3):498–507.

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2010). Individual discount rates and smoking: Evidence from a field experiment in denmark. *Journal of Health Economics*, 29(5):708 – 717.

Harrison, G. W., Lau, M. I., Rutström, E. E., and Sullivan, M. B. (2005c). Eliciting risk and time preferences using field experiments: Some methodological issues. *Field experiments in economics*, 10:125–218.

Harrison, G. W., Lau, M. I., and Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment. *The American Economic Review*, 92(5):1606–1617.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4):1009–1055.

Harrison, G. W., List, J. A., and Towe, C. (2007b). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2):433–458.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2015b). Reduction of compound lotteries with objective probabilities: Theory and evidence. *Journal of Economic Behavior & Organization*, 119:32–55.

Harrison, G. W. and Rutström, E. E. (2008). Risk aversion in the laboratory. In Harrison, G. W. and Cox, J. C., editors, *Risk aversion in experiments*. Emerald Group Publishing, Bingley, UK.

Harrison, G. W. and Swarthout, J. T. (2014). Experimental payment protocols and the bipolar behaviorist. *Theory and Decision*, 77(3):423–438.

Hey, J. D., Morone, A., and Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty*, 39(3):213–235.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, pages 1291–1326.

Hoffmann, A. O., Post, T., and Pennings, J. M. (2013). Individual investor perceptions and behavior during the financial crisis. *Journal of Banking & Finance*, 37(1):60–74.

Holt, C. A. (1986). Preference reversals and the independence axiom. *American Economic Review*, 76:508–515.

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

Holt, C. A. and Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *The American economic review*, 95(3):902–904.

Isaac, R. M. and James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty*, 20(2):177–187.

Jackle, A., Baghal, T. A., Burton, J., and Lynn, P. (2015). User manual. In Jackle, A. and Baghal, T. A., editors, *Understanding Society: Innovation Panel, Waves 1-7*. University of Essex, Colchester.

Jacobson, S. and Petrie, R. (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of Risk and Uncertainty*, 38(2):143–158.

Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., and Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology*, (forthcoming).

Karni, E. (1983). Risk aversion for state-dependent utility functions: Measurement and applications. *International Economic Review*, pages 637–647.

Karni, E. (1987). Generalized expected utility analysis of risk aversion with state-dependent preferences. *International Economic Review*, pages 229–240.

Karni, E. (1993). A definition of subjective probabilities with state-dependent preferences. *Econometrica*, 61(1):187–198.

Karni, E., Schmeidler, D., and Vind, K. (1983). On state dependent preferences and subjective probabilities. *Econometrica*, pages 1021–1031.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., and Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, 8(2):75.

Loomes, G. (1991). Evidence of a new violation of the independence axiom. *Journal of Risk and uncertainty*, 4(1):91–108.

Loomes, G. and Pogrebna, G. (2014). Measuring individual risk attitudes when preferences are imprecise. *The Economic Journal*, 124(576):569–593.

Loomes, G. and Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica*, pages 581–598.

Love, R. O. and Robison, L. J. (1984). An empirical analysis of the intertemporal stability of risk preference. *Southern Journal of Agricultural Economics*, 16(01):159–166.

Lynn, P. (2009). Sample design for understanding society. Understanding Society Working Paper 2009-1, Understanding Society.

March, C., Ziegelmeyer, A., Greiner, B., Cyranek, R., et al. (2016). Pay few subjects but pay them well: Cost-effectiveness of random incentive systems. Technical report, CESifo Group Munich.

Mitchell, S. H. (1999). Measures of impulsivity in cigarette smokers and non-smokers. *Psychopharmacology*, 146(4):455–464.

Mosteller, F. and Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5):371–404.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Reynaud, A. and Couture, S. (2012). Stability of risk preference measures: results from a field experiment on french farmers. *Theory and decision*, 73(2):203–221.

Reynolds, B. (2004). Do high rates of cigarette consumption increase delay discounting?: A cross-sectional comparison of adolescent smokers and young-adult smokers and nonsmokers. *Behavioural processes*, 67(3):545–549.

Reynolds, B. (2006). The experiential discounting task is sensitive to cigarette-smoking status and correlates with a measure of delay discounting. *Behavioural Pharmacology*, 17(2):133–142.

Reynolds, B., Patak, M., Shroff, P., Penfold, R. B., Melanko, S., and Duhig, A. M. (2007). Laboratory and self-report assessments of impulsive behavior in adolescent daily smokers and nonsmokers. *Experimental and clinical psychopharmacology*, 15(3):264–271. ID: 2007-08187-006.

Rieger, M. O., Wang, M., and Hens, T. (2014). Risk preferences around the world. *Management Science*, 61(3):637–648.

Sabater-Grande, G. and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners? dilemma games: An experimental test. *Journal of Economic Behavior & Organization*, 48(1):37–50.

Sahm, C. R. (2012). How much does risk tolerance change? *The Quarterly Journal of Finance*, 2(04):1250020.

Smidts, A. (1997). The relationship between risk attitude and strength of preference: A test of intrinsic risk attitude. *Management Science*, 43(3):357–370.

Starmer, C. and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? an experimental investigation. *The American Economic Review*, 81(4):971–978.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, 103(1):510.

Szrek, H., Chao, L.-W., Ramlagan, S., and Peltzer, K. (2012). Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. *Judgment and decision making*, 7(6):716.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review*, 100(1):557–571.

Vieider, F. M., Chmura, T., Fisher, T., Kusakawa, T., Martinsson, P., Thompson, F. M., and Sunday, A. (2015a). Within-versus between-country differences in risk attitudes: implications for cultural comparisons. *Theory and Decision*, 78(2):209–218.

Vieider, F. M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., and Martinsson, P. (2015b). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, 13(3):421–452.

Viscusi, W. K. and Evans, W. N. (1990). Utility functions that depend on health status: estimates and economic implications. *American Economic Review*, pages 353–374.

Vlaev, I., Chater 1, N., and Stewart, N. (2009). Dimensionality of risk perception: Factors affecting consumer understanding and evaluation of financial risk. *Journal of Behavioral Finance*, 10(3):158–181.

Von Gaudecker, H.-M., Van Soest, A., and Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, pages 664–694.

Wakker, P. and Deneffe, D. (1996). Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. *Management science*, 42(8):1131–1150.

Wakker, P. P. (2008). Explaining the characteristics of the power (crra) utility family. *Health Economics*, 17(12):1329–1344.

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity.* Cambridge University Press.

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

Wehrung, D., MacCrimmon, K., and Brothers, K. (1984). Utility-assessment-domains, stability, and equivalence procedures. *INFORMS Operational Research*, 22(2):98–115.

Wilcox, N. T. (1993). Lottery choice: Incentives, complexity and decision time. *The Economic Journal*, 103(421):1397–1417.

Wölbert, E. and Riedl, A. (2013). Measuring time and risk preferences: Reliability, stability, domain specificity. Discussion paper, CESifo Working Paper Series.

Yi, R., Chase, W. D., and Bickel, W. K. (2007). Probability discounting among cigarette smokers and nonsmokers: molecular analysis discerns group differences. *Behavioral Pharmacology*, 18(7):633–639.

Zeisberger, S., Vrecko, D., and Langer, T. (2012). Measuring the time stability of prospect theory preferences. *Theory and Decision*, 72(3):359–386.

Appendices

A Tables

Table 1: Summary of key studies directly testing cross-validity of different risk preferences measures

Study	**Design**	**Methods**	**Monetary incentives**	**Sample and Setting**	**Cross-measures validity/stability?**
Isaac & James (2000)	Within-subject	BDM, FPA	Real	34 students from University of Arizona, US.	No. Subjects behaved as risk averse in FPA and as risk loving in the BDM. The numerical values of implied CRRA parameters are not stable across the two methods. There is a significant reordering of individuals across methods in terms of the ranking of implied CRRA parameters. The orderings of individual risk tolerance between FPA and BDM are significantly but negatively correlated.
Berg, Dickhaut & McCabe (2005)	Within-subject	BDM, ECA, FPA	Real	48 undergraduate students in Carlson School of Business at University of Minnesota.	No. Estimated risk coefficients are statistically different across BDM, ECA, and FPA methods. Same subjects were risk-seeking in ECA and risk averse in FPA.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Anderson & Mellor (2009)	Within-subject	HL; HRS for job gambles; HRS for inheritance gambles	Real for HL; hypothetical for HRS	236 undergraduate students from College of William and Mary, US.	No. Only for 34% of the subjects the self-reported answers to the HRS questions for job gambles were consistent with their own responses to the HRS questions for inheritance gambles. There was small, marginally significant, correlation between such two HRS measures. For the majority of subjects, the HRS self-reported measures were not significantly associated with the responses to the HL test. There was no monotonic relationship between the level of risk aversion implied by the HRS questions and the range of the CRRA parameter elicited in the HL test.
Hey, Morone & Schmidt (2009)	Within-subject	BDM, MPL, SPA with WTA, SPA with WTP	Real	24 students from the University of York	No. MPL leads to more risk aversion than BDM and SPA. Low, often not significant, and sometimes negative correlation of the estimated CRRA coefficients across the BDM, MPL, and SPA elicitation methods.
Dave, Eckel, Johnson & Rojas (2010)	Within-subject	EG, HL	Real	881 residents in urban and non-urban areas, Canada.	No. There was a significant difference between the risk preferences coefficients estimated using the HL and EG measures. The HL measure generated a significantly higher estimation of CRRA parameter, indicating greater risk aversion, than the EG measure.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Dohmen, Falk, Huffman, Sunde, Schupp & Wagner (2011)	Within-subject	MPL, SOEP-G	Real for MPL. Hypothetical for SOEP	450 subjects from a representative sample, Germany.	Yes. SOEP-G is positive and significantly associated with the value of the safe option at the switching point in the MPL task, although the fraction of variance explained is very low (around 6%).
Levy-Garboua, Maafi, Masclet & Terracol (2012)	Within-subject	HL with 10 simultaneous binary choices; HL with 10 sequential binary choices.	Real	240 students from Universities of Paris 1 and Rennes 1, France.	No. Significantly more inconsistent choices when HL presents 10 binary choices as sequential rather than simultaneous (37% of inconsistent subjects versus 30% of subjects). Significantly higher (about 16-19% higher) CRRA coefficient of risk aversion when HL presents 10 binary choices as sequential rather than simultaneous, at the switching point in the MPL task.
Reynaud & Couture (2012)	Within-subject	DOSPERT, EG, HL	Hypothetical	30 cash crop farmers using irrigation from Midi-Pyrenees, Poitou-Charentes, and Centre region, France.	No. Significantly higher CRRA coefficient of risk aversion with EG than with HL test. Different risk attitudes across different DOSPERT domains. DOSPERT scores in finance and leisure domains are only marginally correlated with EG and HL measures. HL and EG measures are not significantly correlated with DOSPERT scores in general, nor in ethics, health and safety, and social domains.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Szrek, Chao, Ramlagan & Peltzer (2012)	Within-subject	BART, DOSPERT, HL, SOEP-G.	Real for BART and HL. Hypothetical for DOSPERT and SOEP-G.	351 patients of health clinics in Witbank, South Africa.	No. No correlation between SOEP-G and BART. No correlation between SOEP-G and HL. No correlation between BART and DOSPERT. Marginally significant correlation between BART and HL. Marginally significant correlation between BART and DOSPERT. Marginally significant correlation between DOSPERT and HL. Significant correlation between DOSPERT and SOEP-G.
Deck, Lee, Reyes, & Rosen (2013)	Within-subject	BART, DOND, DOSPERT, EG, HL	Real for EG, HL, DOND, and BART. Hypothetical for DOSPERT	203 subjects from University of Arkansas, US.	No. Significant but weak correlation between EG and HL tasks (around 0.2), and between DOND and BART tasks (around 0.27). No significant correlation between any other two tasks. EG and HL tasks are not correlated with DOSPERT scores.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Crosetto & Filippin (2013b)	Between-subject	BART, BRET, CGP, DOSPERT, EG, HL, SOEP-G	Real for BART, BRET, CGP, EG, HL. Hypothetical for DOSPERT, SOEP	444 undergraduate students in Friedrich Schiller University Jena, Germany.	No. Strong risk aversion in EG and HL; moderate risk aversion in BRET and CGP; risk seeking in BART. No correlation between: BART and DOSPERT; BRET and DOSPERT; BRET and SOEP-G; CGP and DOSPERT; CGP and SOEP-G. Marginally significant and weak correlation between HL and SOEP-G (around 0.23). Significant but weak correlation between HL and DOSPERT (around 0.25). Significant correlation between: BART and SOEP-G; EG and DOSPERT; EG and SOEP-G (around 0.30). Significant correlation between DOSPERT and SOEP-G.
Dulleck, Fell, & Fooken (2013)	Within-subject	AH, HL	Real	78 subjects from Queensland University of Technology, Australia.	No. No correlation between AH and HL tasks.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Wolbert and Riedl (2013)	Within-subject	MPL, SOEP-G, SOEP-F, SOEP-H, SOEP in car driving, SOEP in sports/leisure, SOEP in career	Real for MPL. Hypothetical for SOEP.	144 students at Maastricht University, the Netherlands.	Somewhat. Significant but weak correlation between MPL, and SOEP-G and SOEP-F (around 0.2-0.27). No correlation between MPL, and SOEP-H, SOEP in leisure/sports, and SOEP in car driving.
Deck, Lee & Reyes (2014)	Within-subjects	HL gambling task, HL investment task, DOSPERT.	Real for HL. Hypothetical for DOSPERT.	50 students at Chapman University, US.	Somewhat. Higher risk taking in HL gambling task than in HL investment task. DOSPERT attitude towards financial risk is associated with HL investment task. DOSPERT attitude towards gambling is not associated with HL gambling task.
Loomes & Pogrebna (2014)	Within-subjects. Between subjects for HL with ascending or descending order of probability.	AT, EG, HL.	Real.	423 students at University of Warwick.	No. Distribution of responses in HL task depends on whether questions are presented in either ascending or descending order. Significant positive, but often weak, correlations between HL questions. High significant positive correlation between rankings of EG options. Significant positive, but often weak, correlations between AT questions. No correlation between EG and HL tasks. Weak correlation between AT and HL tasks. No correlation between AT and EG tasks.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Charness & Viceisza	Between-subject	CGP, HL, SOEP	Real for CGP and HL, hypothetical for SOEP	91 subjects in Thies and Diourbel, rural Senegal	No. No correlation between CGP and SOEP. No correlation between HL and SOEP.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Vieider, Lefebvre, Bouchouicha, Chmura, Hakimov, Krawczyk & Martinsson (2015)	Within-subjects.	MPL with real rewards in four domains/contexts (risky gains, uncertain gains, risky losses, uncertain losses), SOEP.	Real for MPL. Hypothetical for SOEP.	2,939 students from 30 countries: Australia, Belgium, Brazil, Cambodia, Chile, China, Colombia, Costa Rica, Czech Republic, Ethiopia, France, Germany, Guatemala, India, Japan, Kyrgyzstan, Malaysia, Nicaragua, Nigeria, Peru, Poland, Russia, Saudi Arabia, South Africa, Spain, Thailand, Tunisia, UK, USA, Vietnam.	Somewhat. No significant between-country correlation between MPL with risky gains and SOEP-G. Positive and significant between-country correlation between MPL with risky gains and SOEP-F. For 14 countries out of 30 positive and significant within-country correlation between MPL with risky gains and SOEP-G: no significant correlation between MPL with risky gains and SOEP-G in Belgium, Chile, Ethiopia, France, Guatemala, India, Kyrgyzstan, Nicaragua, Nigeria, Saudi Arabia, South Africa, Spain, Tunisia, UK, US, and Vietnam. For 13 countries out of 30 positive and significant within-country correlation between MPL with risky gains and SOEP-F: no significant correlation between MPL with risky gains and SOEP-F in Australia, Belgium, Cambodia, Chile, Costa Rica, Ethiopia, Guatemala, Japan, Nicaragua, Nigeria, Peru, South Africa, Thailand, Tunisia, UK, US, and Vietnam. No significant correlation between MPL with risky gains and SOEP-H.

Continued on the next page

— Continued

Study	Design	Methods	Monetary incentives	Sample and Setting	Cross-measures validity/stability?
Attanasi, Georgantzis, Rotondi & Vigani (2016)	Within-subject	HL, SG, SOEP-G	Real for HL and SG. Hypothetical for SOEP-G	62 undergraduate students in economics, Bocconi University Milan, Italy.	Somewhat. No significant correlation between HL and SG tasks. Significant correlation between HL and SOEP tasks, and between SG and SOEP tasks.
Falk, Becker, Dohmen, Huffman & Sunde (2016)	Within-subjects.	MPL with real rewards, MPL with hypothetical rewards, SOEP-G.	Real and hypothetical for MPL. Hypothetical for SOEP-G.	409 students at the University of Bonn.	Yes. Significant correlation between MPL with real rewards and MPL with hypothetical rewards (around 0.41). Significant correlation between MPL with real rewards and SOEP-G (around 0.35).
Josef, Richter, Samanez-Larkin, Wagner, Hertwig & Mata (2016)	Within-subjects.	MPL, SOEP-G.	Real for MPL. Hypothetical for SOEP-G.	433 respondents to SOEP survey, Germany.	Yes. Small but significant correlation between MPL and SOEP-G (around 0.24).

Notes - AH: Andreoni & Harbaugh task; AT: Allocation Task; BART: Balloon Analogue Risk Task; BDM: Becker-DeGroot-Marschak procedure; BRET: Bomb Risk Elicitation Task; CGP: Charness-Gneezy-Potters investment game; DOND: Deal or No Deal task; DOSPERT: DOSPERT questionnaire method; ECA: English Clock Auction; EG: Eckel & Grossman method; FPA: First Price Auction method; HL: Holt & Laury method; HRS: Health & Retirement Survey questions; MPL: Multiple Price List method with choices between a safe option and a risky lottery; SG: Sabater-Grande & Georgantzis task; SOEP: German Socio-Economic Panel survey method for self-reported willingness to take risks; SOEP-G: German Socio-Economic Panel survey method for self-reported willingness to take risks in general; SOEP-F: German Socio-Economic Panel survey method for self-reported willingness to take risks in financial matters; SPA: Second Price Auction method.

Table 2: HL lottery probabilities, payoffs, expected values, and relative Sharpe ratios

Low monetary stakes

Lottery A				Lottery B				CRRA Range	
P	GBP	(1-p)	GBP	p	GBP	(1-p)	GBP	Lower bound	Upper bound
0.1	40	0.9	32	0.1	77	0.9	2	$-\infty$	-1.71
0.2	40	0.8	32	0.2	77	0.8	2	-1.71	-0.95
0.3	40	0.7	32	0.3	77	0.7	2	-0.95	-0.49
0.4	40	0.6	32	0.4	77	0.6	2	-0.49	-0.14
0.5	40	0.5	32	0.5	77	0.5	2	-0.14	0.15
0.6	40	0.4	32	0.6	77	0.4	2	0.15	0.41
0.7	40	0.3	32	0.7	77	0.3	2	0.41	0.68
0.8	40	0.2	32	0.8	77	0.2	2	0.68	0.97
0.9	40	0.1	32	0.9	77	0.1	2	0.97	1.37

Low monetary stakes

Lottery A				Lottery B				CRRA Range	
P	GBP	(1-p)	GBP	p	GBP	(1-p)	GBP	Lower bound	Upper bound
0.1	100	0.9	40	0.1	180	0.9	2	$-\infty$	-0.75
0.2	100	0.8	40	0.2	180	0.8	2	-0.75	-0.32
0.3	100	0.7	40	0.3	180	0.7	2	-0.32	-0.05
0.4	100	0.6	40	0.4	180	0.6	2	-0.05	0.16
0.5	100	0.5	40	0.5	180	0.5	2	0.16	0.34
0.6	100	0.4	40	0.6	180	0.4	2	0.34	0.52
0.7	100	0.3	40	0.7	180	0.3	2	0.52	0.7
0.8	100	0.2	40	0.8	180	0.2	2	0.7	0.91
0.9	100	0.1	40	0.9	180	0.1	2	0.91	1.2

Table 3: Choices in the EG task

	Payoff		**Expected**	**Standard**	**CRRA Ranges**	
	Low	**High**	**Return**	**Deviation**	Lower bound	Upper bound
A	28	28	28	0	$-\infty$	0
B	24	36	30	8.5	0	0.499
C	20	44	32	17	0.499	0.71
D	16	52	34	25.5	0.71	1.16
E	12	60	36	33.9	1.16	3.46
F	2	70	36	48.1	3.46	$+\infty$

Table 4: Descriptive statistics

	All respondents				Repeated Respondents			
	IP6		IP7		IP6		IP7	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Dependent Children	0.017	0.006	0.008	0.006	0.019	0.010	0.011	0.008
Female	0.528	0.021	0.519	0.023	0.539	0.031	0.522	0.032
Age	49.47	1.000	50.02	1.245	47.53	1.368	48.94	1.336
Unemployed	0.039	0.008	0.028	0.012	0.047	0.015	0.024	0.011
Self Employed	0.079	0.012	0.079	0.016	0.065	0.016	0.074	0.018
Employee	0.456	0.023	0.505	0.026	0.530	0.028	0.556	0.027
Married	0.436	0.022	0.440	0.031	0.487	0.042	0.484	0.041
ALevel	0.417	0.027	0.439	0.031	0.476	0.036	0.478	0.037
Per Capita Income	1966	176	1499	49	2124	311	1550	60
Web	0.331	0.022	0.391	0.030	0.382	0.038	0.378	0.039
N	661		453		410		400	

	IP6			IP7		
	Mean	St.Dev.	N	Mean	St.Dev.	N
Smoker	0.197	0.021	608	0.173	0.024	467
BMI	26.31	0.239	611			
Junk Food	2.515	0.055	608			
Fruit and Vegetables				2.20	0.049	432
Heavy Drinking	2.329	0.082	488	2.472	0.104	344
Savings	166.21	33.44	573			
Savings Horizon	1.981	0.052	342			
Regular Saver	4.291	1.391	342			
Personal Pension	0.097	0.012	657			
DOSPERT F				5.720	1.018	467
DOSPERT H				7.827	1.066	467

Notes - Descriptive statistics computed using sampling weights. Standard errors adjusted to account for strata and PSU in the survey.

Table 5: Number of subjects for each "switching point" in the HL tasks with low and high monetary stakes in IP6 and IP7.

	Low Stakes					
	IP6			**IP7**		
First Choice of *Lottery B*	All	Consistent	Inconsistent	All	Consistent	Inconsistent
1	218	116	102	160	75	85
2	36	13	23	42	7	35
3	30	17	13	22	9	13
4	49	41	8	25	19	6
5	62	55	7	42	37	5
6	42	39	3	27	22	5
7	37	35	2	22	20	2
8	20	18	2	27	26	1
9	23	23		21	21	
Never	138	138		80	80	
Total	655	495	160	468	316	152

	High Stakes					
	IP6			**IP7**		
First Choice of *Lottery B*	All	Consistent	Inconsistent	All	Consistent	Inconsistent
1	225	153	72	167	101	66
2	37	14	23	28	10	18
3	21	12	9	23	11	12
4	41	31	10	28	22	6
5	84	76	8	47	42	5
6	47	44	3	32	30	2
7	35	35	1	25	24	1
8	22	21		19	19	
9	15	15		13	13	
Never	127	127		86	86	
Total	654	528	126	468	358	110

Table 6: Number of safe choices in the HL tasks (IP6)

	Low Stakes				**High Stakes**			
	All Subjects		**Consistent**		**All Subjects**		**Consistent**	
Number of *Lottery A*	Total	Fraction	Total	Fraction	Total	Fraction	Total	Fraction
0	116	17.71	116	23.43	153	23.39	153	28.98
1	25	3.82	14	2.83	26	3.98	14	2.65
2	36	5.5	17	3.43	27	4.13	12	2.27
3	68	10.38	41	8.28	50	7.65	31	5.87
4	68	10.38	55	11.11	93	14.22	76	14.39
5	71	10.84	39	7.88	70	10.7	44	8.33
6	53	8.09	35	7.07	55	8.41	35	6.63
7	35	5.34	18	3.64	30	4.59	21	3.98
8	45	6.87	22	4.44	23	3.52	15	2.84
9	138	21.07	138	27.88	127	19.42	127	24.05
Total	655		495		654		528	

Table 7: Choices in the EG task

	All Subjects		CRRA Ranges	
	Total	**Fraction**	Lower bound	Upper bound
A	230	35.11	$-\infty$	0
B	64	9.77	0	0.499
C	120	18.32	0.499	0.71
D	84	12.82	0.71	1.16
E	53	8.09	1.16	3.46
F	104	15.88	3.46	$+\infty$
Total	655			

Table 8: Responses to SOEP Questions

	General		**Financial**		**Health**	
Choice	**Total**	**Fraction**	**Total**	**Fraction**	**Total**	**Fraction**
Missing	26	3.93	27	4.08	26	3.93
Unwilling	22	3.33	68	10.29	84	12.71
1	46	6.96	98	14.83	77	11.65
2	71	10.74	110	16.64	123	18.61
3	86	13.01	84	12.71	89	13.46
4	69	10.44	65	9.83	57	8.62
5	113	17.1	82	12.41	67	10.14
6	50	7.56	28	4.24	37	5.6
7	70	10.59	44	6.66	39	5.9
8	49	7.41	16	2.42	21	3.18
9	11	1.66	3	0.45	2	0.3
Fully Prepared	16	2.42	4	0.61	6	0.91
Total	629		629		628	

Notes - 32 respondents were classified as inapplicable.

Figure 1: Distribution of responses (IP6)

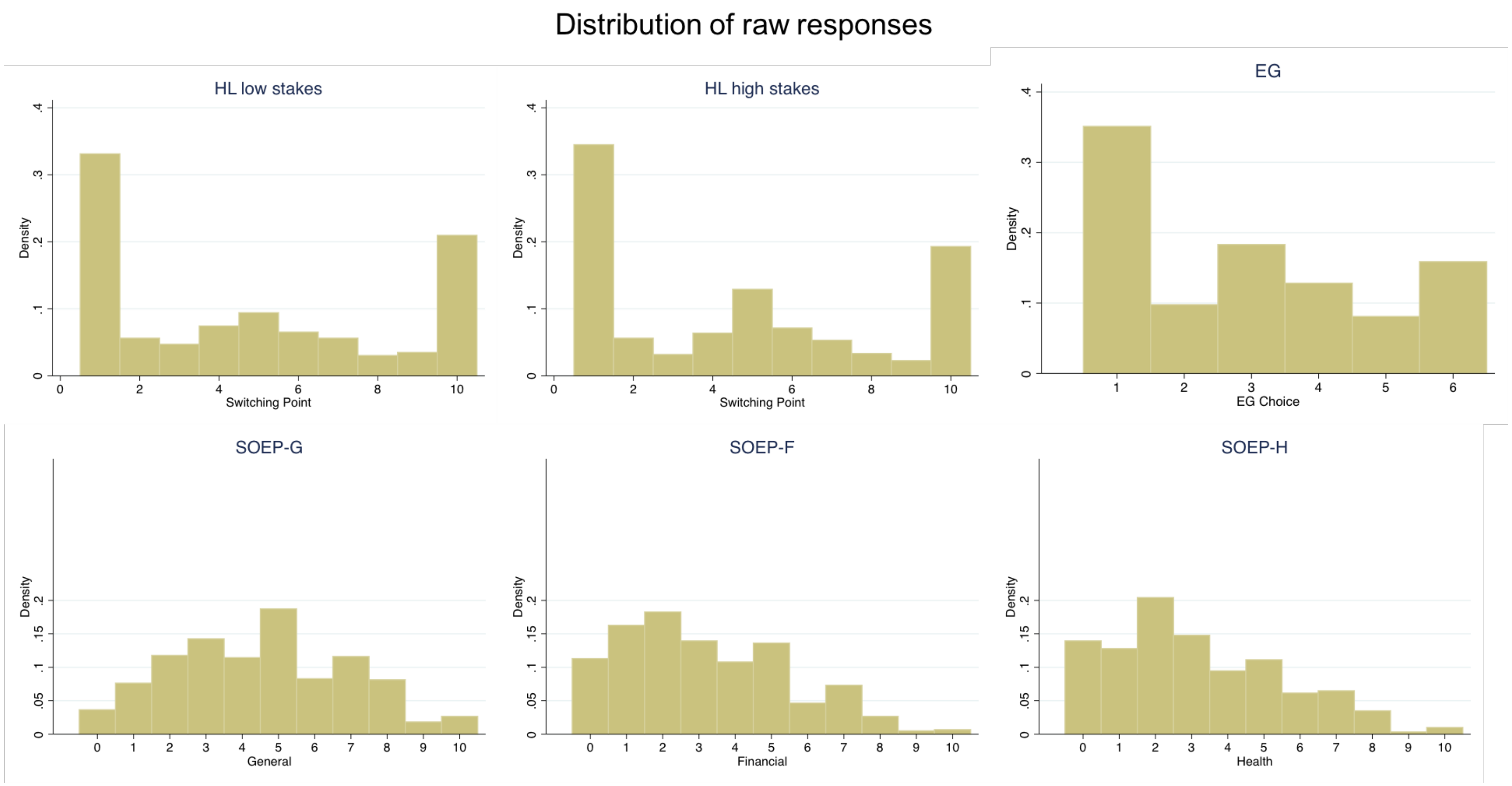

Table 9: Cross-methods pairwise correlations in IP6 and IP7

IP6

	EG RA	**HL RA**		**Survey**	
		Low Stakes	High Stakes	SOEP-G RA	SOEP-F RA
HL RA					
Low Stakes	**0.120**				
	(0.009)				
	[655]				
High Stakes	**0.172**	**0.669**			
	(0.000)	(0.000)			
	[654]	[654]			
Survey					
SOEP-G RA	**0.095**	0.057	**0.014**		
	(0.058)	(0.249)	(0.013)		
	[606]	[606]	[606]		
SOEP-F RA	**0.136**	0.191	0.077	**0.621**	
	(0.007)	(0.719)	(0.190)	(0.000)	
	[605]	[605]	[605]	[608]	
SOEP-H RA	**0.103**	-0.018	-0.00003	**0.531**	**0.565**
	(0.015)	(0.692)	(0.999)	(0.000)	(0.000)
	[606]	[606]	[606]	[606]	[606]

IP7

	EG RA	**HL RA**		**Survey**	
		Low Stakes	High Stakes	SOEP-G RA	SOEP-F RA
HL RA					
Low Stakes	**0.129**				
	(0.000)				
	[467]				
High Stakes	**0.169**	**0.627**			
	(0.004)	(0.000)			
	[467]	[467]			
Survey					
SOEP-G RA	0.007	0.055	-0.055		
	(0.892)	(0.359)	(0.313)		
	[433]	[433]	[433]		
SOEP-F RA	0.019	0.0399	-0.007	**0.4989**	
	(0.752)	(0.558)	(0.905)	(0.000)	
	[432]	[432]	[432]	[432]	
SOEP-H RA	-0.014	-0.056	**-0.125**	**0.420**	**0.439**
	(0.825)	(0.340)	(0.052)	(0.000)	(0.000)
	[432]	[432]	[432]	[432]	[432]

Notes - Number of observations in square brackets; p-value in brackets. All correlations computed using sampling weights. Standard errors account for strata and PSU. All measures are increasing in risk aversion.

Table 10: Cross-methods pairwise correlations for respondents in both IP6 and IP7

	EG RA		**HL RA**		
Lotteries	IP6	IP7	IP6		IP7
			Low Stakes	High Stakes	Low Stakes
EG RA					
IP7	**0.138**				
	(0.028)				
	[407]				
HL RA					
IP6 Low Stakes	0.061	**0.196**			
	(0.357)	(0.009)			
	[407]	[407]			
IP6 High Stakes	**0.131**	0.089	**0.694**		
	(0.010)	(0.198)	(0.000)		
	[407]	[407]	[407]		
IP7 Low Stakes	0.009	**0.139**	**0.225**	**0.203**	
	(0.893)	(0.034)	(0.003)	(0.007)	
	[407]	[410]	[407]	[407]	
IP7 High Stakes	0.024	**0.137**	**0.224**	**0.219**	**0.609**
	(0.730)	(0.043)	(0.007)	(0.008)	(0.000)
	[407]	[410]	[407]	[407]	[410]

	SOEP-G RA		**SOEP-F RA**		**SOEP-H RA**
Survey	IP6	IP7	IP6	IP7	IP6
SOEP-G RA IP7	**0.551**				
	(0.000)				
	[367]				
SOEP-F RA IP6		**0.358**			
		(0.000)			
		[366]			
SOEP-F RA IP7	**0.4595**		**0.473**		
	(0.000)		(0.000)		
	[366]		[365]		
SOEP-H RA IP6		**0.3298**		**0.262**	
		(0.000)		(0.000)	
		[366]		[365]	
SOEP-H RA IP7	**0.292**		**0.241**		**0.447**
	(0.000)		(0.000)		(0.000)
	[366]		[365]		[365]

Notes - Number of observations in square brackets; p-value in brackets. All correlations computed using sampling weights. Standard errors account for strata and PSU. All measures are increasing in risk aversion.

Table 11: EG RA in IP6 and IP7

	IP7						
IP6	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
(-∞, 0]	12.0	11.6	14.9	23.7	2.5	35.4	59
(0,0.499]	10.6	15.7	13.7	15.6	6.7	37.8	31
(0.499, 0.71]	14.9	18.7	21.9	19.1	8.0	17.3	56
(0.71, 1.16]	4.7	14.3	18.9	26.6	10.5	25.0	84
(1.16, 3.46]	3.0	0.0	3.4	18.0	20.3	55.4	31
(3.46, +∞)	16.5	7.8	3.4	17.6	8.5	46.3	145
Total	11.7	11.2	11.6	20.4	8.7	36.3	406

HL IP6 Consistent

	IP7						
IP6	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
(-∞, 0]	16.0	6.5	9.9	19.0	3.7	44.9	37
(0,0.499]	14.5	13.0	18.7	14.9	4.1	34.9	25
(0.499, 0.71]	20.9	13.6	17.6	22.3	5.3	20.4	30
(0.71, 1.16]	4.6	8.5	19.3	28.1	13.6	26.0	64
(1.16, 3.46]	4.8	0.0	0.0	20.5	22.9	51.8	19
(3.46, +∞)	11.3	10.3	2.9	16.2	8.3	51.0	108
Total	11.3	9.3	10.3	20.1	9.2	40.0	283

Notes - Percentage of individuals EG RA x in IP7, by EG RA in IP6. Individuals are defined as *HL IP6 Consistent* if they have less than 2 switching points in the HL task, in IP6.

Table 12: HL RA with low monetary stakes in IP6 and IP7

HL IP6 All respondents

Panel A	**IP7**										
IP6	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
(-∞,-1.71]	52.96	10.86	4.40	2.98	2.07	3.85	1.02	0.81	3.20	17.85	130
(-1.71,-0.95]	43.36	13.53	15.66	3.09	9.50	0	9.07	2.89	0	2.89	26
(-0.95,-0.49]	21.96	16.33	0	19.36	14.14	8.32	0	0	5.77	14.12	21
(-0.49,-0.14]	25.93	7.89	0	12.90	13.28	4.09	0	17.88	2.61	15.42	32
(-0.14,0.15]	27.27	18.83	2.52	4.13	17.02	6.38	0	12.80	3.38	7.67	37
(0.15,0.41]	15.21	4.84	3.37	22.47	17.85	14.71	0	12.33	2.92	6.30	37
(0.41,0.68]	37.19	5.23	0	12.45	3.99	11.48	19.72	9.94	0	0	23
(0.68,0.97]	24.20	0	0	10.01	0	0	18.74	0	12.51	34.55	9
(0.97,1.37]	28.76	0	5.17	16.51	0	4.32	12.46	23.28	0	9.50	20
(1.37,+∞)	39.57	0.78	2.22	0	5.80	1.44	3.11	5.32	3.89	37.87	72
Total	37.81	8.39	3.58	7.29	7.45	5.01	3.58	6.77	3.06	17.05	407

HL IP6 Consistent

Panel B	**IP7**										
IP6	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
(-∞,-1.71]	47.86	4.63	7.11	5.92	4.12	7.63	0	0	6.36	16.37	64
(-1.71,-0.95]	33.10	0	44.82	0	12.09	0	0	0	0	9.99	7
(-0.95,-0.49]	0	18.27	0	29.64	21.65	0	0	0	8.83	21.61	13
(-0.49,-0.14]	24.14	0	0	15.39	12.71	4.89	0	21.34	3.12	18.41	26
(-0.14,0.15]	25.02	20.14	3.05	4.99	17.95	0	0	15.48	4.09	9.27	30
(0.15,0.41]	15.14	2.34	3.77	25.17	20	16.47	0	13.80	0	3.30	32
(0.41,0.68]	38.61	5.42	0	12.92	4.15	8.11	20.47	10.32	0	0	22
(0.68,0.97]	0	0	0	16.49	0	0	30.87	0	20.60	32.04	5
(0.97,1.37]	18.42	0	0	21.81	0	0	16.46	30.76	0	12.55	15
(1.37,+∞)	38.44	0.81	2.31	0	6.05	1.50	3.25	5.55	4.05	38.03	68
Total	31.50	4.93	4.07	9.99	9.39	5.04	3.78	8.91	3.94	18.43	282

Table 13: HL RA with high monetary stakes in IP6 and IP7

HL IP6 All respondents

Panel A	**IP7**										
IP6	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
(-∞,-0.75]	54.23	4.75	6.14	2.07	4.79	3.68	0.79	2.36	0	21.19	133
(-0.75,-0.32]	18.96	13.07	17.56	6.31	7.02	12.62	0	0	4.88	19.59	22
(-0.32,-0.05]	55.72	0	24.31	0	0	3.99	0	0	3.94	12.04	16
(-0.05,0.16]	45.46	0	3.80	14.82	12.89	19.40	3.63	0	0	0	26
(0.16,0.34]	31.77	4.33	6.95	11.53	18.39	9.65	7.69	2.18	0	7.51	64
(0.34,0.52]	29.54	4.78	5.19	6.05	14.48	10.67	24.34	4.95	0	0	25
(0.52,0.70]	24.21	0	3.42	12.38	12.93	5.01	4.92	0	5.49	31.64	32
(0.70,0.91]	38.48	6.52	0	4.43	6.66	4.14	14.17	0	0	25.61	14
(0.91,1.2]	0	0	0	0	39.18	0	0	18.90	28.15	13.77	6
(1.2,+∞)	32.26	2.49	1.94	3.35	3.45	1.48	9.85	3.90	5.49	35.79	68
Total	39.33	3.88	6.18	5.84	8.99	6.25	5.72	2.37	2.21	19.22	406

HL IP6 Consistent

Panel B	**IP7**										
IP6	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
(-∞,-0.75]	49.71	5.78	2.34	3.43	7.92	4.95	1.31	2.75	0	21.82	77
(-0.75,-0.32]	16.46	0	0	24.31	0	22.93	0	0	0	36.30	6
(-0.32,-0.05]	18.91	0	57.97	0	0	0	0	0	9.40	13.73	7
(-0.05,0.16]	44.57	0	0	12.00	13.71	24.41	5.31	0	0	0	18
(0.16,0.34]	24.45	4.01	8.97	13.32	17.74	9.09	9.93	2.81	0	9.68	51
(0.34,0.52]	18.76	0	6.42	7.49	17.91	13.19	30.11	6.12	0	0	18
(0.52,0.70]	21.21	0	0	14.18	14.81	5.74	5.64	0	6.29	32.14	25
(0.70,0.91]	43.07	0	0	4.95	7.45	0	15.86	0	0	28.67	13
(0.91,1.2]	0	0	0	0	39.18	0	0	18.90	28.15	13.77	6
(1.2,+∞)	28.41	2.77	2.16	3.73	3.84	1.65	10.96	4.34	6.11	36.03	62
Total	32.86	2.91	4.53	7.35	10.73	6.69	8.06	3.02	2.74	21.12	283

Table 14: SOEP-G RA in IP6 and IP7

All respondents

	IP7											
SOEP-G RA	0	1	2	3	4	5	6	7	8	9	10	Total
0	46.50	22.16	8.39	0	12.99	9.97	0	0	0	0	0	15
1	37.89	0	0	41.08	21.03	0	0	0	0	0	0	4
2	0	15.23	19.51	19.61	4.14	8.10	4.82	24.58	0	4.01	0	25
3	3.27	5.16	14.51	42.87	5.29	11.63	1.90	10.73	4.62	0	0	48
4	13.14	0	0	18.63	12.38	24.64	5.44	18.85	0	6.92	0	29
5	1.10	0	4.41	12.17	11.09	40.22	6.26	8.99	9.21	5.21	1.34	63
6	2.10	0	0	12.77	11.54	32.55	10.23	12.64	13.91	2.27	1.98	45
7	0	0	3.71	3.63	5.51	14.38	16.91	11.51	27.21	13.26	3.88	47
8	0	0	7.69	9.70	3.01	10.75	7.19	5.45	21.24	26.39	8.58	50
9	0	8.56	0	0	0	24.38	14.83	4.92	21.94	10.25	15.13	23
10	0	0	0	0	12.31	0	0	13.50	31.85	14.16	28.18	6
Total	4.47	3.27	6.04	14.80	7.63	20.80	7.63	10.74	12.54	8.42	3.66	355

Table 15: SOEP-F RA in IP6 and IP7

All respondents

	IP7											
SOEP-F RA	0	1	2	3	4	5	6	7	8	9	10	Total
0	0	0	49.89	0	0	0	0	0	0	0	50.11	3
1	60.87	0	39.13	0	0	0	0	0	0	0	0	2
2	18.25	0	13.57	39.92	0	0	0	11.22	17.05	0	0	9
3	0	11.14	5.54	21.87	7.78	25.49	16.90	3.49	7.80	0	0	26
4	0	0	0	6.95	21.61	32.69	0	4.10	14.22	0	20.42	19
5	0	0	3.16	2.28	21.69	28.77	10.30	6.84	16.61	6.58	3.76	53
6	0	0	0	0	11.88	16.25	0	20.76	24.13	11.82	15.16	37
7	0	0	1.24	0	7.33	15.47	12.55	22.88	16.69	10.57	13.25	55
8	2.44	0	0	0	4.38	12.19	4.11	10.61	20.82	27.36	18.09	66
9	0	2.45	2.29	0	1.64	7.28	5.79	8.42	17.77	28.15	26.23	53
10	8.10	0	0	0	2.54	7.58	0	0	0	39.09	42.70	32
Total	2.06	1.18	2.45	3.29	8.61	16.02	6.35	10.70	15.85	16.67	16.83	355

Table 16: SOEP-H RA in IP6 and IP7

All respondents

	IP7											
SOEP-H RA	0	1	2	3	4	5	6	7	8	9	10	Total
0	0	0	63.83	0	36.17	0	0	0	0	0	0	5
1												0
2	0	11.75	30.46	0	7.98	26.19	0	0	6.51	7.33	9.77	14
3	3.73	8.83	3.24	12.65	0	32.33	4.82	14.65	7.52	12.24	0	27
4	16.62	0	0	11.96	2.84	39.34	10.59	0	4.07	10.68	3.90	23
5	0	0	5.53	10.46	0	19.33	5.73	20.80	17.95	13.69	6.52	38
6	3.78	0	0	4.00	2.34	13.30	7.25	18.98	35.89	2.12	12.35	42
7	5.52	1.62	9.73	0	10.41	9.12	4.18	26.59	10.05	10.07	12.71	57
8	0	0	1.95	2.31	1.22	7.68	4.91	17.21	23.42	7.40	33.91	67
9	0	0	3.50	0	0	6.74	0	7.55	21.88	40.86	19.47	39
10	4.51	1.80	0	0	5.07	6.25	0	2.17	13.33	25.96	40.91	42
Total	3.25	1.61	5.31	3.81	3.83	14.16	4.15	14.26	17.33	14.23	18.05	354

Table 17: EG RA and SOEP-G RA in IP6 and IP7

SOEP-G RA	EG RA						
	IP6						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	14.4	5.36	12.27	11.86	10.65	45.46	18
1	31.85	7.47	0	29.41	0	31.27	11
2	13.88	7.64	12.82	34.84	5.65	25.15	44
3	23.02	7.83	8.17	20.31	12.84	27.83	65
4	22.8	8.36	8.6	13.82	8.09	38.33	45
5	15.05	13.81	19.63	12.93	9.13	29.45	117
6	17.82	7.45	12.86	19.62	8.99	33.26	72
7	12.23	9.13	17.08	16.37	5.51	39.67	88
8	8.19	8.74	14.01	14.75	16.85	37.46	71
9	16.79	7.44	3.98	15.89	9.89	46.01	46
10	14.39	6.09	10.37	12.37	11.88	44.9	22
Total	15.91	9.1	13.08	17.4	9.62	34.9	600

SOEP-G RA	IP7						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	8.59	14.95	32.85	5.08	15.03	23.49	18
1	9.73	14.12	21.05	35.5	0	19.59	13
2	9.24	7.09	6.58	24.31	11.75	41.03	29
3	10.01	10.38	9.98	25.83	7.64	36.15	47
4	2.84	12.88	12.15	10.93	22.42	38.78	38
5	15.05	8.72	9.39	28.51	4.65	33.69	92
6	6.58	12.99	14.12	25.06	12.27	28.97	34
7	12.52	16.52	5.68	22.56	7.49	35.23	44
8	16.53	10.98	10.15	13.22	6.83	42.28	67
9	12.98	11.28	18.2	23.78	10.57	23.2	32
10	9.02	0	22.93	10.99	18.07	38.98	16
Total	11.53	10.92	12.09	21.35	9.49	34.62	430

Table 18: EG RA and SOEP-F RA in IP6 and IP7

SOEP-F RA	EG RA						
	IP6						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	61.65	0	0	38.35	0	0	18
1	57.7	0	0	42.3	0	0	11
2	24.34	13.13	0	16.06	0	46.47	44
3	18.72	5.04	9.38	34.33	12.18	20.34	65
4	18.9	17.34	29.08	19.78	0	14.9	45
5	19.79	6.75	11.65	12.86	13.64	35.31	117
6	25.04	9.59	13.23	13.09	2.13	36.93	72
7	12.22	10.25	16.17	18.41	10.6	32.36	88
8	11.3	8.61	13.21	12.82	15.54	38.53	71
9	12.81	11.73	13.49	21.31	5.33	35.33	46
10	10.64	4.54	11.39	14.23	12.57	46.63	22
Total	15.94	8.89	13.11	17.44	9.64	34.97 599	

SOEP-F RA	EG RA						
	IP7						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	18.46	29.69	0	0	19.29	32.57	9
1	47.06	40.8	0	12.14	0	0	3
2	24.85	0	6.49	37.91	0	30.75	12
3	22.59	18.97	18.52	20.61	3.75	15.55	17
4	4.87	9.97	17.37	15.44	15.71	36.63	29
5	8.25	7.51	13.76	26.78	6.17	37.53	82
6	5.82	4.12	7.88	35.23	9.38	37.56	18
7	5.67	15.85	9.29	17.5	5.84	45.85	36
8	13.01	15.9	4.36	18.71	8.49	39.52	72
9	10.53	14.98	4.77	21.92	19.52	28.27	70
10	14.45	2.91	25.49	18.94	6.45	31.76	80
Total	11.55	10.94	12.11	21.39	9.51	34.5	429

Table 19: EG RA and SOEP-H RA in IP6 and IP7

SOEP-H RA	EG RA						
	IP6						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	0	0	0	43.25	56.75	0	5
1	0	0	100	0	0	0	1
2	10.77	0	0	35.84	1.39	52	16
3	7.95	15.58	11.58	36.05	9.22	19.62	35
4	34.35	8.17	9.99	18.02	7.32	22.15	45
5	20.42	10.36	16.76	9.81	3.14	39.52	66
6	18.02	4.31	13.03	21.69	7.6	35.35	62
7	16.98	14.57	17.98	14.62	11.32	24.54	91
8	11.18	5.69	12.46	15.24	12.44	42.99	116
9	18	16.91	9.65	13.7	5.82	35.93	76
10	9.91	3.74	13.47	16.65	14.03	42.2	86
Total	15.91	9.1	13.08	17.4	9.62	34.9	600

SOEP-H RA	EG RA						
	IP7						
	(-∞, 0]	(0,0.499]	(0.499, 0.71]	(0.71, 1.16]	(1.16, 3.46]	(3.46, +∞)	Total
0	23.92	13.81	27.35	9.88	15.9	9.14	18
1	0	7.57	0	34.28	0	58.15	13
2	16.87	16.27	15.27	15.76	0	35.84	29
3	0	16.17	21.22	8.69	14.44	39.49	47
4	0	9.89	19.44	26.96	0	43.71	38
5	15.11	7.84	2.21	29.83	7.83	37.19	92
6	0	9.39	10.87	18.81	16.57	44.36	34
7	4.04	12.21	10.2	29.25	8.7	35.61	44
8	16.79	15.02	11.72	15.81	13.74	26.92	67
9	12.52	7.38	12.92	16.84	11.56	38.77	32
10	11.91	8.76	16.48	22.95	6.67	33.22	16
Total	11.55	10.94	12.11	21.39	9.51	34.5	430

Table 20: HL RA and SOEP-G RA in IP6

SOEP-G RA	**HL - First Switching Point Low Stakes**										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	43.53	13.53	0	2.68	0	16.82	0	0	12.55	10.89	18
1	61.42	0	0	0	7.47	0	9.72	0	0	21.38	11
2	27.83	0	0	13.31	15.26	9.73	4.55	2.97	13.18	13.15	44
3	26.81	7.89	7.34	8.13	11.53	8.25	9.05	1.56	4.72	14.72	65
4	31.25	9.04	6.77	4.19	9.69	4.86	5.43	10.94	3.68	14.14	45
5	41.61	8.6	1.74	9.09	5.19	6.45	3.88	2.47	2.49	18.47	117
6	41.35	11.89	1.36	6.18	4.59	4.49	5.63	0.85	6.62	17.02	72
7	28.25	0	13.46	13.09	14.73	0.78	4.57	0.99	2.63	21.51	88
8	28.31	3.72	7.99	5.95	9.95	13.13	4.9	1.98	0.82	23.25	71
9	21.04	5.68	4.54	5.62	12.57	5.84	6.25	0	1.89	36.56	46
10	49.15	5.62	0	5.84	0	4.75	2.53	9.64	0	22.47	22
Total	33.73	6.14	5.09	8.05	9.12	6.59	5.16	2.53	4.06	19.54	600

SOEP-G RA	**HL - First Switching Point High Stakes**										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	60.36	14.21	0	7.28	0	0	0	7.26	0	10.89	18
1	53.39	19.2	0	0	7.47	9.72	0	0	10.21	0	11
2	26.7	6.9	4.73	4.16	14.03	5.54	25.21	4.84	0	7.9	44
3	30.63	2.7	2.53	13.67	17.03	10.46	6.3	2.11	2.42	12.15	65
4	31.02	7.33	0	17.83	10.99	5.26	4.73	7.45	6	9.38	45
5	39.24	9.74	6.71	3.43	15.18	7.27	4.58	0.7	0	13.16	117
6	47.77	5.18	0	4.18	7.11	6.39	0	8.42	2.26	18.69	72
7	28.68	4.05	6.05	3.47	22.19	6.8	3.07	0	5.32	20.38	88
8	26.07	2.85	4.2	6.83	10.68	7.51	7.78	1.71	2.56	29.81	71
9	21.52	2.51	8.96	2.22	17.08	6.25	6.51	1.89	0	33.06	46
10	42.26	18.35	0	0	4.12	4.75	2.53	0	0	27.98	22
Total	34.3	6.42	3.99	6.04	13.65	6.84	5.76	2.88	2.26	17.86	600

Table 21: HL RA and SOEP-F RA in IP6

SOEP-F RA	HL - First Switching Point Low Stakes										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	61.65	38.35	0	0	0	0	0	0	0	0	3
1	0	57.7	0	0	0	0	0	0	42.3	0	2
2	43.47	0	0	0	15.21	0	11.21	7.3	0	22.81	18
3	29.7	1.49	4.09	6.07	8.68	18.38	6.49	0	12.16	12.94	45
4	29.09	4.43	11.09	7.55	16.46	8.29	0	0	8.42	14.66	26
5	32.26	6.65	1.56	11.25	8.75	4.82	9.12	3.9	7.63	14.06	85
6	38.43	1.57	4.02	11.91	8.39	1.7	7.55	3.46	0	22.98	62
7	39.59	6.94	8.09	9.87	5.8	12.02	4.15	0	1.22	12.33	82
8	29.08	12.04	1.91	8.52	8.72	6.52	2.23	3.59	4.16	23.24	108
9	30.79	3.21	12.44	5.04	9.92	4.82	6.45	1.02	3.26	23.04	100
10	38.04	5.98	1.38	6.47	10.46	1.51	2.03	5.09	0.97	28.06	69
Total	33.81	6.16	5.1	8.07	9.14	6.38	5.17	2.53	4.06	19.58	599

SOEP-F RA	HL - First Switching Point High Stakes										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	61.65	0	0	38.35	0	0	0	0	0	0	3
1	0	57.7	0	0	0	0	0	42.3	0	0	2
2	43.47	0	0	0	15.21	0	18.51	0	0	22.81	18
3	38.73	3.63	2.54	0.78	12.2	4.95	23.38	0	7.26	6.52	45
4	37.48	0	1.57	12.66	21.59	6.1	8.13	0	4.46	8	24
5	35.82	10.03	7.5	8.58	11.69	7.73	3.57	4.73	0	10.36	85
6	36.13	6.51	2.43	7.26	8.53	13.06	2.07	1.58	6.9	15.55	62
7	41.16	2.64	4.33	6.78	13.17	5.13	4.74	4.99	1	16.06	82
8	25.04	7.51	6.24	4.85	16.42	5.41	4.63	2.45	2.15	25.3	108
9	30.41	4.9	4.24	7.71	8.32	11.55	3.66	3.26	1.83	24.12	101
10	36.74	11.59	0	1.67	23.75	1.51	0.8	2.03	0	21.9	69
Total	34.38	6.44	4	6.05	13.68	6.86	5.54	2.88	2.27	17.9	599

Table 22: HL RA and SOEP-H RA in IP6

SOEP-H RA	**HL - First Switching Point Low Stakes**										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	0	26.55	0	9.77	0	0	0	0	45.74	17.95	5
1	100	0	0	0	0	0	0	0	0	0	1
2	39.9	0	0	26.86	7.3	2.34	0	0	0	23.6	16
3	31.52	1.91	13.43	2.45	8.59	7.01	8.35	3.77	8.93	14.04	35
4	37.08	3.68	8.92	4.16	12.55	4.54	7.98	5.66	4.59	10.84	45
5	29.03	5.46	5.68	9.21	4.58	2.17	7.46	3.08	7.09	26.24	66
6	43.1	1.57	4.03	12.31	8.18	5.73	4.73	0.56	2.56	17.23	62
7	37.35	9.49	1.44	6.44	5.2	10.84	4.53	1.06	2.5	21.14	91
8	28.56	7.47	5.22	5.53	10.41	8.72	4.23	2.71	3.95	23.21	116
9	31.09	2.25	6.29	11.32	10.92	5.52	6.66	1.69	1.24	23.02	76
10	35.27	11.08	3.97	7.06	13.4	6.32	2.88	4.04	3.24	12.72	86
Total	33.73	6.14	5.09	8.05	9.12	6.59	5.16	2.53	4.06	19.54	600

SOEP-H RA	**HL - First Switching Point High Stakes**										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	0	29.03	0	26.55	0	0	0	26.47	0	17.95	5
1	100	0	0	0	0	0	0	0	0	0	1
2	38.05	0	4.19	4.19	20.89	9.08	0	0	6.64	16.96	16
3	29.17	5.03	0	10.25	21.78	0	17.76	4.04	0	11.97	35
4	25.87	8.54	9.23	15.82	15.6	10.02	7.41	0	0	7.5	46
5	41.46	9.28	2.2	3.35	3.3	7.4	7.6	1.56	2.62	21.21	64
6	37.89	6.33	2.43	6.52	14.32	8.59	4.93	0	1.32	17.66	62
7	40.89	4.43	5.22	0.38	11.98	8.62	4.49	3.42	4.95	15.62	92
8	26.56	0.99	4.42	7.48	16.35	5.29	5.31	4.65	1.41	27.55	117
9	31.84	4.52	7.01	6.53	8.28	8.79	6.61	1.86	3.63	20.91	77
10	39.07	14.9	0.96	3.7	18.93	4.82	1.99	4.08	1.21	10.36	87
Total	34.3	6.42	3.99	6.04	13.65	6.84	5.76	2.88	2.26	17.86	600

Table 23: HL RA and SOEP-G RA - Consistent

SOEP-G RA	**HL - First Switching Point Low Stakes**										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	39.41	10.44	0	3.84	0	24.12	0	0	6.57	15.61	13
1	66.08	0	0	0	9.25	0	12.04	0	0	12.63	9
2	19.64	0	0	14.82	17	10.84	5.07	3.31	14.68	14.65	40
3	25.86	1.95	2.66	10.7	11.59	8.4	11.9	1.35	6.22	19.37	50
4	16.02	3.51	9.91	6.13	14.18	7.11	7.94	11.8	5.39	18.01	31
5	27.83	3.01	1.5	13.7	4.9	11	5.74	1.26	3.17	27.9	69
6	33.97	3.03	2.17	5.55	7.33	7.17	7.24	1.36	10.57	21.6	46
7	20.19	0	6.59	13.68	17.64	1.2	1.65	1.54	4.08	33.42	57
8	21.81	1.03	5.95	5.22	10.99	15.39	6.5	2.63	1.09	29.39	54
9	13.1	4.82	3.52	7.97	13.73	2.33	8.86	0	2.69	42.99	33
10	26.2	0	0	0	0	10.36	5.52	8.98	0	48.95	10
Total	24.69	2.19	3.55	9.46	10.95	8.64	6.48	2.52	5.42	26.11	412

SOEP-G RA	**HL - First Switching Point High Stakes**										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	39.41	10.44	0	3.84	0	24.12	0	0	6.57	15.61	13
1	66.08	0	0	0	9.25	0	12.04	0	0	12.63	9
2	19.64	0	0	14.82	17	10.84	5.07	3.31	14.68	14.65	40
3	25.86	1.95	2.66	10.7	11.59	8.4	11.9	1.35	6.22	19.37	51
4	16.02	3.51	9.91	6.13	14.18	7.11	7.94	11.8	5.39	18.01	31
5	27.83	3.01	1.5	13.7	4.9	11	5.74	1.26	3.17	27.9	67
6	33.97	3.03	2.17	5.55	7.33	7.17	7.24	1.36	10.57	21.6	46
7	20.19	0	6.59	13.68	17.64	1.2	1.65	1.54	4.08	33.42	58
8	21.81	1.03	5.95	5.22	10.99	15.39	6.5	2.63	1.09	29.39	55
9	13.1	4.82	3.52	7.97	13.73	2.33	8.86	0	2.69	42.99	33
10	26.2	0	0	0	0	10.36	5.52	8.98	0	48.95	10
Total	24.69	2.19	3.55	9.46	10.95	8.64	6.48	2.52	5.42	26.11	412

Table 24: HL RA and SOEP-F RA - Consistent

SOEP-F RA	**HL - First Switching Point Low Stakes**										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	61.65	38.35	0	0	0	0	0	0	0	0	4
1	0	0	0	0	0	0	0	0	100	0	1
2	43.47	0	0	0	15.21	0	11.21	7.3	0	22.81	18
3	24.97	0	0	7.04	10.06	21.31	7.52	0	14.1	15	39
4	24.61	0	11.23	10.79	8.53	11.84	0	0	12.04	20.95	18
5	18.93	2.06	2.52	15.86	12.94	7.8	13.23	0	8.14	18.51	53
6	22.37	2.31	5.92	12.53	8.11	2.51	11.12	5.1	0	30.03	42
7	27.98	4.51	7.59	12.61	7.06	15.73	6.2	0	1.82	16.51	55
8	23.64	1.27	2.69	8.71	12.09	7.77	3.14	5.06	5.86	29.75	77
9	22.67	2.05	4.1	7.59	10.19	4.85	4.64	1.13	5.46	37.33	60
10	24.7	2.08	0	5.36	16.73	2.42	3.24	5.35	1.56	38.56	43
Total	24.77	2.2	3.56	9.49	10.98	8.34	6.5	2.53	5.44	26.19	411

SOEP-F RA	**HL - First Switching Point High Stakes**										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	61.65	0	0	38.35	0	0	0	0	0	0	4
1	0	0	0	0	0	0	0	100	0	0	1
2	43.47	0	0	0	15.21	0	18.51	0	0	22.81	18
3	36.53	0	2.16	0.9	11.57	5.74	27.11	0	8.42	7.56	39
4	31.35	0	2.31	12.1	21.83	8.95	5.18	0	6.55	11.74	17
5	29.74	3.12	4.14	9.42	13.89	9.52	5.77	7.64	0	16.75	53
6	34.36	0	3.58	6.14	10.11	14.24	3.04	2.32	6.44	19.77	43
7	33.06	2.1	2.72	7.08	16.33	7.67	3.75	5.64	1.49	20.17	55
8	27.12	1.8	4.21	4.27	15.8	7.62	5.74	2.6	3.03	27.81	78
9	23.57	1.93	1.91	8.12	8.42	11.61	2.43	5.46	3.06	33.48	61
10	32.44	1.83	0	0	26.57	2.42	1.29	3.25	0	32.21	43
Total	31.07	1.51	2.65	5.72	14.73	8.06	6.74	3.84	2.96	22.73	411

Table 25: HL RA and SOEP-H RA - Consistent

SOEP-H RA	**HL - First Switching Point** **Low Stakes**										
	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
0	0	37.41	0	13.77	0	0	0	0	23.53	25.29	4
1	100	0	0	0	0	0	0	0	0	0	1
2	31.4	0	0	30.66	8.33	2.67	0	0	0	26.94	14
3	26.64	0	5.21	3.34	11.7	5.31	11.38	5.14	12.16	19.12	26
4	25.04	0	10.66	6.6	14.22	7.2	12.65	4.46	7.29	11.89	29
5	25.82	2.68	0	11.49	4.78	3.52	7.51	0	9.68	34.53	41
6	24.06	2.18	5.57	17.05	11.32	7.93	6.56	0	3.55	21.79	45
7	26.03	4.24	0	7.99	6.41	17.01	3.96	1.67	3.92	28.78	58
8	19.52	0.65	5	6.03	12.94	10.43	4.82	3.68	5.37	31.55	86
9	22.45	1.25	4.31	10.88	10.34	5.61	9.51	2.42	1.77	31.46	54
10	30.73	3.51	1.93	5.91	17.18	8.91	4.67	4.3	5.25	17.61	54
Total	24.69	2.19	3.55	9.46	10.95	8.64	6.48	2.52	5.42	26.11	411

SOEP-H RA	**HL - First Switching Point** **High Stakes**										
	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
0	0	0	0	37.41	0	0	0	37.3	0	25.29	4
1	100	0	0	0	0	0	0	0	0	0	1
2	31.4	0	2.67	4.78	23.84	10.37	0	0	7.58	19.35	14
3	30.58	0	0	4.08	19.33	0	24.2	5.5	0	16.31	26
4	29.77	0	0	16.92	13.77	15.9	11.75	0	0	11.89	29
5	39.76	4.29	3.66	1.21	0	5.11	12.61	2.59	4.34	26.43	39
6	30.73	2.56	3.37	9.03	15.43	8.73	6.83	0	1.82	21.51	45
7	33.91	2.35	2.72	0.59	14.98	11.87	2.97	5.34	5.07	20.19	59
8	24.74	0	1.12	7.51	15.86	5.21	7.22	4.39	1.91	32.04	87
9	25.53	2.16	7.61	3.68	9.07	10.9	5.29	2.66	5.19	27.9	54
10	38.57	1.47	1.56	3.84	25.13	6.62	1.04	6.61	1.96	13.2	54
Total	30.97	1.5	2.64	5.7	14.69	8.03	7.05	3.82	2.95	22.65	411

Table 26: EG RA and HL RA in IP6

	HL RA - First Switching Point Low Stakes										
EG RA	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
(-∞, 0]	50.55	6.22	0	10.63	8.55	5.92	4.2	3.07	0	10.85	108
(0,0.499]	36.83	0	12.65	8.26	11.34	6.8	5.8	1.38	0	16.95	61
(0.499, 0.71]	30.41	11.23	5.56	7.31	14.18	5.75	8.33	1.56	5.24	10.44	84
(0.71, 1.16]	22.91	5.86	3.06	10.84	7.91	10.75	9.17	3.42	7.28	18.81	116
(1.16, 3.46]	28.43	11.58	3.42	8.48	8.24	6.68	3.42	4.07	10.35	15.33	63
(3.46, +∞)	32.98	4.24	6.8	5.73	7.48	6.26	2.58	2.29	2.98	28.66	223
Total	33.69	6.06	5.09	8.14	9.03	7.02	5.13	2.61	3.97	19.27	655

	HL RA - First Switching Point High Stakes										
EG RA	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
(-∞, 0]	56.09	7.31	3.79	3.59	10.97	5.92	1.42	0	2.85	8.06	108
(0,0.499]	35.98	10.89	1.86	7.09	17.13	1.16	8.46	1.3	2.55	13.59	62
(0.499, 0.71]	33.42	4.56	5.11	8.73	23.65	5.99	6.75	2.62	2.03	7.14	82
(0.71, 1.16]	22.73	4.3	5.21	4.46	16.54	11.38	11.4	3.92	2.52	17.53	116
(1.16, 3.46]	25.07	10.03	6.92	5.87	14.25	4.78	0	9.55	6.63	16.9	63
(3.46, +∞)	33.26	4.37	1.83	6.52	9.58	6.11	5.29	3.48	0.98	28.57	223
Total	34.65	6.03	3.66	5.94	13.97	6.4	5.7	3.26	2.39	18	654

Table 27: EG RA and HL RA in IP6 - Consistent

	HL RA - First Switching Point Low Stakes										
EG RA	(-∞,-1.71]	(-1.71,-0.95]	(-0.95,-0.49]	(-0.49,-0.14]	(-0.14,0.15]	(0.15,0.41]	(0.41,0.68]	(0.68,0.97]	(0.97,1.37]	(1.37,+∞)	Total
(-∞, 0]	46.81	4.14	0	13.76	9.91	8.17	4.17	0	0	13.03	79
(0,0.499]	18.9	0	11.62	7.65	15	11.2	7.96	0	0	27.94	37
(0.499, 0.71]	14.56	2.07	5.37	9.23	20	10.67	13.71	2.13	6.08	15.83	46
(0.71, 1.16]	12.9	1.64	1.21	12.24	10	14.36	10.76	3.37	10.31	23.07	82
(1.16, 3.46]	13.5	1.6	3.8	10.19	13	5.99	6.18	7.35	14.51	23.81	35
(3.46, +∞)	25.49	2.79	3.92	6.81	8	6.91	3.41	3.03	3.49	35.95	170
Total	24.33	2.42	3.52	9.6	11	9.16	6.53	2.55	4.96	25.91	449

	HL RA - First Switching Point High Stakes										
EG RA	(-∞,-0.75]	(-0.75,-0.32]	(-0.32,-0.05]	(-0.05,0.16]	(0.16,0.34]	(0.34,0.52]	(0.52,0.70]	(0.70,0.91]	(0.91,1.2]	(1.2,+∞)	Total
(-∞, 0]	57.13	2.77	4.85	2.59	11.3	8.17	1.96	0	3.93	7.3	79
(0,0.499]	23.36	4.41	3.06	5.64	25	1.91	13.94	0	0	22.41	38
(0.499, 0.71]	18.28	2.18	6.73	6.7	31	11.37	10.23	2.29	3.86	7.72	44
(0.71, 1.16]	17.1	0	1.49	6.33	19	13.56	13.51	5.55	3.57	19.9	82
(1.16, 3.46]	18.77	0	2.74	2.98	23	4.16	0	12.52	11.99	23.76	35
(3.46, +∞)	32	1.36	0.43	5.91	8	5.32	5.52	4.6	1.3	35.11	170
Total	30.59	1.59	2.42	5.23	16	7.55	7.09	3.98	3.16	22.8	448

Table 28: CRRA estimation

IP6

	Interval Regression					
	All			Consistent		
	HL low RA	HL high RA	EG RA	HL low RA	HL high RA	EG RA
RA estimate	-0.60 (0.12)***	-0.12 (0.089)	1.98 (0.14)***	0.038 (0.14)	0.13 (0.11)	2.18 (0.19)***
Observations	655	655	655	449	449	449

IP7

	Interval Regression					
	All			Consistent		
	HL low RA	HL high RA	EG RA	HL low RA	HL high RA	EG RA
RA estimate	-0.84 (0.18)***	-0.18 (0.099)*	2.13 (0.18)***	-0.038 (0.22)	0.10 (0.13)	2.25 (0.26)***
Observations	467	467	467	280	280	280

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 29: CRRA estimation with controls - Consistent

	Interval Regression					
	IP6			IP7		
	HL low RA	HL high RA	EG RA	HL low RA	HL high RA	EG RA
RA estimate	-0.56	-0.41	4.76	-3.85	-2.83	-3.29
	(1.25)	(0.96)	(1.98)**	(2.41)	(1.34)**	(2.36)
Female	0.12	0.087	1.00	-0.26	-0.22	1.15
	(0.25)	(0.17)	(0.43)**	(0.47)	(0.29)	(0.42)***
Age	-0.0039	-0.0021	0.013	0.0021	0.0028	0.0037
	(0.0082)	(0.0055)	(0.012)	(0.014)	(0.0095)	(0.013)
Married	-0.086	-0.14	0.19	-0.13	-0.15	-0.39
	(0.27)	(0.20)	(0.36)	(0.42)	(0.24)	(0.38)
A Level	-0.16	-0.046	-0.040	0.0012	0.055	-0.90
	(0.31)	(0.20)	(0.40)	(0.38)	(0.25)	(0.43)**
Per Capita Income (ln)	0.14	0.11	-0.57	0.54	0.41	0.66
	(0.19)	(0.14)	(0.27)**	(0.32)*	(0.17)**	(0.34)*
Self Employed	-0.84	-0.72	-0.33	-0.15	0.090	1.55
	(0.52)	(0.35)**	(0.83)	(0.89)	(0.55)	(0.79)*
Unemployed	0.15	0.31	-0.060	0.87	0.86	2.42
	(0.76)	(0.48)	(0.81)	(1.71)	(1.05)	(1.65)
Employee	-0.23	-0.24	0.49	-0.069	-0.00064	0.99
	(0.28)	(0.18)	(0.44)	(0.46)	(0.26)	(0.42)**
Web (interview)	0.054	0.082	0.22			
	(0.28)	(0.19)	(0.36)			
Observations	448	448	448	280	280	280

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode. An F-test for the joint significance of all the coefficients fails to reject the H_0 of risk neutrality for both the HL tasks.

Table 30: Cross-validity ordered probit estimations (IP6)

	All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL low RA estimate	0.031 (0.022)	0.017 (0.024)	-0.0072 (0.020)			
HL high RA estimate				0.11 (0.041)**	0.060 (0.046)	-0.0099 (0.040)
Observations	609	608	608	609	608	608

	Consistent					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL low RA estimate	0.066 (0.027)**	0.053 (0.031)*	0.0034 (0.022)			
HL high RA estimate				0.15 (0.048)***	0.093 (0.051)*	0.00036 (0.039)
Observations	420	419	419	420	419	419

	All			Consistent		
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
EG RA estimate	0.039 (0.020)*	0.059 (0.021)***	0.044 (0.016)**	0.057 (0.024)**	0.069 (0.022)***	0.066 (0.021)***
Observations	609	608	608	420	419	419

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 31: Cross-validity ordered probit estimations: HL low stakes with controls (IP6)

	All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL low RA estimate	0.026	0.0079	-0.014	0.035	0.012	-0.012
	(0.021)	(0.024)	(0.019)	(0.022)	(0.024)	(0.020)
EG RA estimate	0.037	0.058	0.045	0.029	0.050	0.036
	(0.020)*	(0.020)***	(0.016)***	(0.021)	(0.020)**	(0.017)**
Female				0.33	0.25	0.16
				(0.11)***	(0.096)**	(0.10)
Age				0.017	0.012	0.012
				(0.0037)***	(0.0031)***	(0.0038)***
Married				-0.053	-0.0025	0.11
				(0.11)	(0.11)	(0.10)
A Level				-0.049	-0.16	0.032
				(0.11)	(0.11)	(0.097)
Per Capita Income (ln)				-0.075	-0.082	-0.20
				(0.077)	(0.074)	(0.061)***
Self Employed				0.099	-0.24	-0.15
				(0.18)	(0.15)	(0.15)
Unemployed				0.12	0.033	0.025
				(0.24)	(0.33)	(0.24)
Employee				-0.19	-0.13	-0.037
				(0.11)*	(0.12)	(0.11)
Web (interview)				-0.013	0.042	0.10
				(0.11)	(0.11)	(0.11)
Observations	609	608	608	607	606	606

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 32: Cross-validity ordered probit estimations: HL low stakes with controls - Consistent (IP6)

	All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL low RA estimate	0.058	0.042	-0.0086	0.060	0.044	-0.0053
	(0.026)**	(0.030)	(0.023)	(0.030)**	(0.031)	(0.024)
EG RA estimate	0.050	0.064	0.068	0.035	0.046	0.052
	(0.024)**	(0.022)***	(0.022)***	(0.024)	(0.022)**	(0.023)**
Female				0.34	0.31	0.11
				(0.13)**	(0.12)**	(0.13)
Age				0.019	0.014	0.013
				(0.0038)***	(0.0034)***	(0.0037)***
Married				-0.17	-0.024	0.12
				(0.14)	(0.13)	(0.12)
A Level				0.035	-0.15	-0.050
				(0.15)	(0.13)	(0.11)
Per Capita Income (ln)				-0.15	-0.15	-0.23
				(0.081)*	(0.084)*	(0.070)***
Self Employed				0.056	-0.049	-0.040
				(0.22)	(0.20)	(0.17)
Unemployed				0.45	0.045	0.029
				(0.34)	(0.38)	(0.28)
Employee				-0.20	0.091	0.085
				(0.15)	(0.14)	(0.15)
Web (interview)				0.079	0.11	0.12
				(0.12)	(0.11)	(0.12)
Observations	420	419	419	419	418	418

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 33: Cross-validity ordered probit estimations: HL high stakes with controls (IP6)

	All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL high RA estimate	0.098	0.042	-0.026	0.13	0.059	-0.016
	(0.041)**	(0.045)	(0.041)	(0.040)***	(0.043)	(0.040)
EG RA estimate	0.030	0.054	0.046	0.019	0.045	0.036
	(0.021)	(0.020)***	(0.018)**	(0.021)	(0.020)**	(0.018)*
Female				0.35	0.26	0.16
				(0.11)***	(0.094)***	(0.10)
Age				0.018	0.012	0.012
				(0.0036)***	(0.0031)***	(0.0039)***
Married				-0.052	-0.0012	0.11
				(0.12)	(0.12)	(0.11)
A Level				-0.064	-0.16	0.033
				(0.11)	(0.11)	(0.097)
Per Capita Income (ln)				-0.081	-0.085	-0.20
				(0.075)	(0.073)	(0.061)***
Self Employed				0.14	-0.22	-0.15
				(0.18)	(0.15)	(0.15)
Unemployed				0.096	0.025	0.031
				(0.24)	(0.33)	(0.24)
Employee				-0.16	-0.12	-0.039
				(0.11)	(0.12)	(0.11)
Web (interview)				-0.021	0.036	0.10
				(0.11)	(0.11)	(0.11)
Observations	609	608	608	607	606	606

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 34: Cross-validity ordered probit estimations: HL high stakes with controls - Consistent (IP6)

	All					
	SOEP-H RA	SOEP-F	SOEP-H	SOEP-H RA	SOEP-F	SOEP-H
HL high RA estimate	0.14	0.071	-0.025	0.15	0.084	-0.011
	(0.048)***	(0.051)	(0.041)	(0.049)***	(0.051)	(0.041)
EG RA estimate	0.044	0.062	0.069	0.026	0.043	0.053
	(0.025)*	(0.023)***	(0.023)***	(0.025)	(0.023)*	(0.024)**
Female				0.35	0.32	0.11
				(0.13)***	(0.12)**	(0.13)
Age				0.019	0.014	0.013
				(0.0039)***	(0.0034)***	(0.0037)***
Married				-0.16	-0.017	0.12
				(0.14)	(0.13)	(0.12)
A Level				0.027	-0.15	-0.049
				(0.15)	(0.13)	(0.11)
Per Capita Income (ln)				-0.16	-0.16	-0.23
				(0.076)**	(0.081)*	(0.069)***
Self Employed				0.10	-0.032	-0.043
				(0.23)	(0.20)	(0.17)
Unemployed				0.42	0.027	0.032
				(0.34)	(0.38)	(0.28)
Employee				-0.18	0.100	0.084
				(0.15)	(0.14)	(0.15)
Web (interview)				0.070	0.100	0.12
				(0.12)	(0.11)	(0.12)
Observations	420	419	419	419	418	418

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 35: Cross-validity estimations - HL low RA ranges (IP6)

	Interval Regression, HL low RA ranges							
	All				Consistent			
EG RA estimate	0.14				0.15			
	(0.047)***				(0.047)***			
SOEP-G RA		0.073				0.13		
		(0.051)				(0.059)**		
SOEP-F RA			0.037				0.11	
			(0.060)				(0.068)	
SOEP-H RA				-0.010				0.022
				(0.047)				(0.049)
Observations	655	606	605	606	449	417	416	417

	Interval Regression, HL low RA ranges - with controls							
	All				Consistent			
EG RA estimate	0.14				0.15			
	(0.048)***				(0.049)***			
SOEP-G RA		0.098				0.14		
		(0.059)				(0.071)*		
SOEP-F RA			0.046				0.11	
			(0.064)				(0.074)	
SOEP-H RA				-0.0080				0.022
				(0.051)				(0.054)
Observations	653	604	603	604	448	416	415	416
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 36: Cross-validity estimations - HL high RA ranges (IP6)

	Interval Regression, HL high RA ranges							
	All				Consistent			
EG RA estimate	0.12 (0.028)***				0.12 (0.035)***			
SOEP-G RA		0.090 (0.035)**				0.13 (0.045)***		
SOEP-F RA			0.060 (0.042)				0.089 (0.050)*	
SOEP-H RA				0.004 (0.033)				0.011 (0.036)
Observations	655	606	605	606	449	417	416	417

	Interval Regression, HL high RA ranges - with controls							
	All				Consistent			
EG RA estimate	0.12 (0.027)***				0.13 (0.036)***			
SOEP-G RA		0.12 (0.036)***				0.14 (0.049)***		
SOEP-F RA			0.078 (0.041)*				0.099 (0.052)*	
SOEP-H RA				0.008 (0.033)				0.018 (0.037)
Observations	653	604	603	604	448	416	415	416
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 37: Cross-validity estimations - EG RA ranges (IP6)

	Interval Regression, EG RA ranges									
	All					Consistent				
HL low RA estimate	0.18 (0.062)***					0.24 (0.080)***				
HL high RA estimate		0.40 (0.091)***					0.47 (0.13)***			
SOEP-G RA			0.12 (0.062)*					0.18 (0.083)**		
SOEP-F RA				0.18 (0.064)***					0.23 (0.081)***	
SOEP-H RA					0.13 (0.049)**					0.22 (0.071)***
Observations	655	655	606	605	606	449	449	417	416	417

	Interval Regression, EG RA ranges									
	All					Consistent				
HL low RA estimate	0.18 (0.064)***					0.25 (0.081)***				
HL high RA estimate		0.43 (0.089)***					0.49 (0.13)***			
SOEP-G RA			0.10 (0.067)					0.14 (0.087)		
SOEP-F RA				0.16 (0.066)**					0.18 (0.086)**	
SOEP-H RA					0.10 (0.053)*					0.18 (0.076)**
Observations	653	653	604	603	604	448	448	416	415	416
Controls	x	x	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 38: B-EG RA temporal stability across IP6 and IP7 - interval regression

Dependent variable: EG RA IP7, repeated individuals				
	All		**Consistent**	
EG RA estimate IP6	0.14	0.16	0.18	0.17
	(0.066)**	(0.068)**	(0.090)*	(0.090)*
Female		0.52		0.97
		(0.32)		(0.40)**
Age		0.018		0.023
		(0.011)		(0.015)
Married		-0.64		-0.56
		(0.37)*		(0.51)
A Level		-0.16		-0.74
		(0.33)		(0.42)*
Per Capita Income (ln)		0.41		0.46
		(0.34)		(0.49)
Self Employed		-0.020		0.13
		(0.65)		(0.92)
Unemployed		-0.023		13.8
		(1.23)		(0.97)***
Employee		-0.11		-0.042
		(0.43)		(0.53)
Web (interview)		0.12		-0.45
		(0.39)		(0.44)
Constant	1.90	-1.89	2.07	-2.01
	(0.22)***	(2.59)	(0.30)***	(3.66)
Observations	410	399	287	278

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 39: HL RA temporal stability across IP6 and IP7 - interval regression

Dependent variable: HL RA IP7, repeated individuals

	Low stakes				**High stakes**			
	All		**Consistent**		**All**		**Consistent**	
HL low RA estimate	0.26	0.27	0.20	0.21				
	(0.085)***	(0.081)***	(0.089)**	(0.090)**				
HL high RA estimate					0.28	0.32	0.26	0.28
					(0.11)***	(0.095)***	(0.10)**	(0.10)***
Female		-0.48		-0.13		-0.62		-0.48
		(0.33)		(0.37)		(0.21)***		(0.27)*
Age		-0.000042		0.0022		0.016		0.017
		(0.012)		(0.015)		(0.0086)*		(0.0099)
Married		0.25		-0.21		-0.23		-0.56
		(0.33)		(0.38)		(0.23)		(0.28)**
A Level		-0.17		-0.033		0.26		0.22
		(0.35)		(0.39)		(0.24)		(0.27)
Per Capita Income (ln)		0.17		0.24		-0.22		-0.096
		(0.33)		(0.39)		(0.25)		(0.26)
Self Employed		-0.70		-0.33		-0.50		-0.26
		(0.65)		(0.66)		(0.46)		(0.52)
Unemployed		0.25		-2.87		1.94		2.30
		(2.00)		(1.27)**		(1.25)		(0.55)***
Employee		-0.81		-0.48		-0.18		-0.10
		(0.41)*		(0.38)		(0.29)		(0.29)
Web (interview)		0.25		0.16		0.048		-0.13
		(0.35)		(0.44)		(0.26)		(0.27)
Constant	-0.71	-1.33	-0.47	-1.86	-0.22	0.97	-0.029	0.39
	(0.20)***	(2.56)	(0.19)**	(3.05)	(0.12)*	(1.87)	(0.13)	(2.11)
Observations	410	399	287	278	410	399	287	278

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 40: SOEP RA temporal stability across IP6 and IP7 - ordered probit

Dependent variable: SOEP RA IP7, repeated individuals

	SOEP-G		**SOEP-F**		**SOEP-H**	
	All	**Consistent**	**All**	**Consistent**	**All**	**Consistent**
SOEP-G	0.28	0.26				
	(0.039)***	(0.041)***				
SOEP-F			0.24	0.25		
			(0.034)***	(0.029)***		
SOEP-H					0.21	0.19
					(0.039)***	(0.037)***
Female		0.39		0.41		0.31
		(0.12)***		(0.15)***		(0.14)**
Age		0.0077		0.0067		0.012
		(0.0060)		(0.0030)**		(0.0044)***
Married		-0.082		0.085		0.040
		(0.14)		(0.17)		(0.20)
A Level		0.22		0.066		0.0048
		(0.16)		(0.15)		(0.16)
Per Capita Income (ln)		0.021		0.22		0.19
		(0.14)		(0.16)		(0.15)
Self Employed		0.32		0.14		1.05
		(0.33)		(0.24)		(0.39)***
Unemployed		-0.30		1.72		0.049
		(1.02)		(0.62)***		(0.91)
Employee		-0.031		-0.15		0.089
		(0.20)		(0.18)		(0.19)
Web (interview)		-0.047		-0.16		0.11
		(0.15)		(0.20)		(0.16)
Observations	367	360	365	358	365	358

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Figure 2: Estimation of midpoints - HL low stakes (IP6)

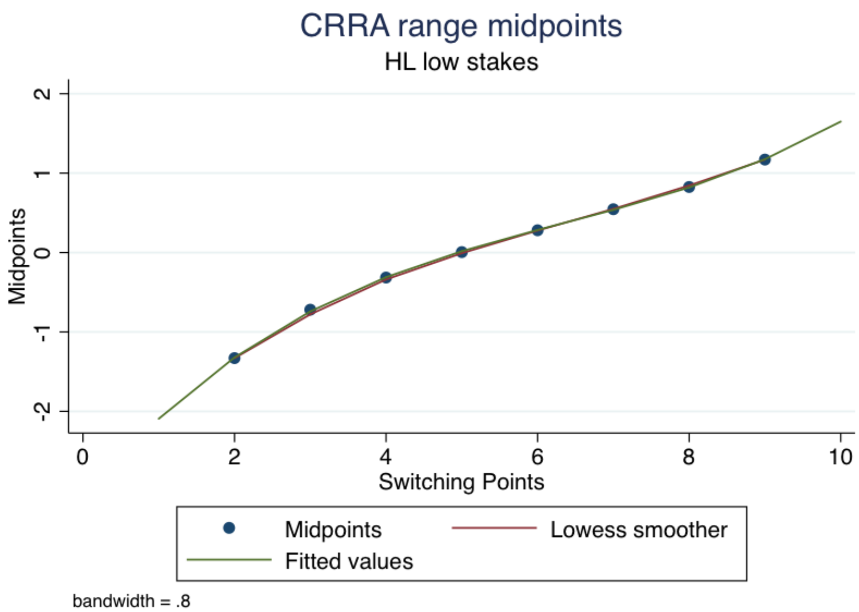

Figure 3: Estimation of midpoints - EG RA (IP6)

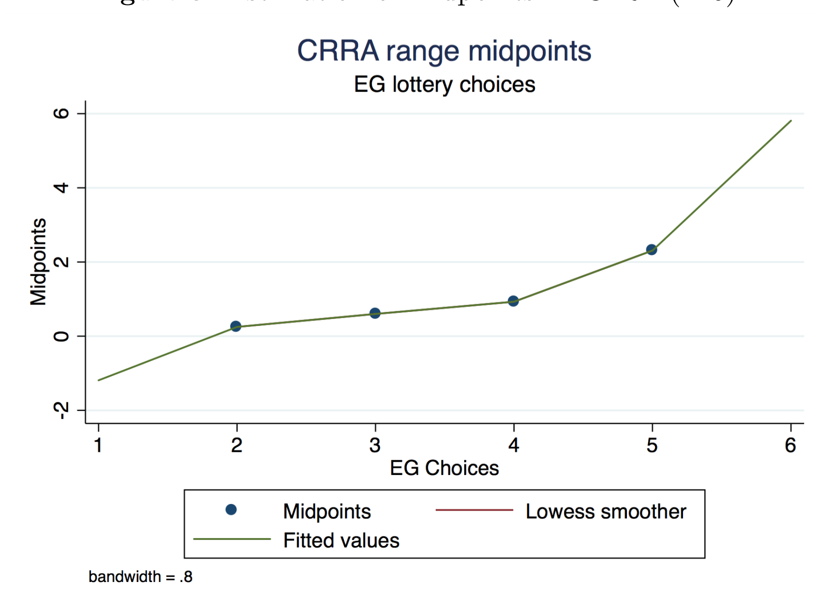

Table 41: Predicting smoking status using RA measures (IP6)

	Dependent variable: Smoker ==1 if smokes cigarettes					
HL low RA estimate	-0.029	-0.024				
	(0.043)	(0.044)				
HL high RA estimate			-0.030	-0.028		
			(0.058)	(0.059)		
EG RA estimate					-0.023	-0.018
					(0.032)	(0.031)
Controls		x		x		x
Observations	608	606	608	606	608	606

	Dependent variable: Smoker ==1 if smokes cigarettes									
SOEP-G RA	-0.043	-0.044					-0.040	-0.042	-0.040	-0.042
	(0.029)	(0.032)					(0.030)	(0.033)	(0.030)	(0.033)
SOEP-F RA			-0.032	-0.032						
			(0.034)	(0.037)						
SOEP-H RA					-0.013	-0.00023				
					(0.029)	(0.033)				
HL low RA estimate							-0.025	-0.020		
							(0.042)	(0.044)		
EG RA estimate							-0.017	-0.013	-0.019	-0.014
							(0.031)	(0.030)	(0.031)	(0.031)
HL high RA estimate									-0.013	-0.0096
									(0.056)	(0.058)
Controls		x		x		x		x		x
Observations	608	606	607	605	607	605	608	606	608	606

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 42: Predicting BMI using RA measures (IP6)

	Dependent variable: Body Mass Index (BMI)					
HL low RA estimate	0.20	0.22				
	(0.13)	(0.13)				
HL high RA estimate			0.42	0.48		
			(0.18)**	(0.19)**		
EG RA estimate					0.19	0.20
					(0.090)**	(0.090)**
Controls		x		x		x
Observations	611	610	611	610	611	610

	Dependent variable: Body Mass Index (BMI									
SOEP-G RA	0.021	-0.062					-0.0039	-0.090	-0.020	-0.12
	(0.096)	(0.10)					(0.096)	(0.10)	(0.097)	(0.10)
SOEP-F RA			0.036	-0.023						
			(0.12)	(0.13)						
SOEP-H RA					0.086	0.050				
					(0.15)	(0.16)				
HL low RA estimate							0.15	0.18		
							(0.14)	(0.14)		
EG RA estimate							0.21	0.21	0.19	0.19
							(0.095)**	(0.096)**	(0.096)**	(0.097)*
HL high RA estimate									0.33	0.42
									(0.19)*	(0.21)**
Controls		x		x		x		x		x
Observations	565	564	564	563	564	563	565	564	565	564

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 43: Predicting junk food consumption using RA measures (IP6)

	Dependent variable: Days eating junk food - Ordered Probit					
HL low RA estimate	-0.021	-0.026				
	(0.025)	(0.025)				
HL high RA estimate			-0.017	-0.020		
			(0.041)	(0.042)		
EG RA estimate					-0.0030	-0.0067
					(0.023)	(0.024)
Controls		x		x		x
Observations	608	606	608	606	608	606

	Dependent variable: Days eating junk food - Ordered Probit									
SOEP-G RA	0.042	-0.013					0.044	-0.010	0.045	-0.011
	(0.025)*	(0.028)					(0.025)*	(0.028)	(0.025)*	(0.028)
SOEP-F RA			0.061	0.016						
			(0.024)**	(0.028)						
SOEP-H RA					0.030	-0.0053				
					(0.020)	(0.022)				
HL low RA estimate							-0.024	-0.025		
							(0.025)	(0.025)		
EG RA estimate							-0.0047	-0.0031	-0.0048	-0.0042
							(0.025)	(0.025)	(0.025)	(0.025)
HL high RA estimate									-0.027	-0.015
									(0.040)	(0.043)
Controls		x		x		x		x		x
Observations	608	606	607	605	607	605	608	606	608	606

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 44: Predicting fruit and vegetables consumption using RA measures (IP7) - Ordered Probit

	Dependent variable: fruit and vegetables portions per day (categories)					
HL low RA estimate	0.059 (0.034)*	0.054 (0.030)*				
HL high RA estimate			0.12 (0.055)**	0.081 (0.042)*		
EG RA estimate					0.017 (0.027)	0.024 (0.029)
Controls		x		x		x
Observations	432	419	432	419	432	419

	Dependent variable: fruit and vegetables portions per day (categories)									
SOEP-G RA	-0.016 (0.024)	0.027 (0.028)					-0.020 (0.024)	0.024 (0.028)	-0.013 (0.025)	0.028 (0.028)
SOEP-F RA			-0.061 (0.028)**	-0.025 (0.030)						
SOEP-H RA					-0.043 (0.024)*	0.0075 (0.024)				
HL low RA estimate							0.059 (0.035)*	0.050 (0.030)		
EG RA estimate							0.0092 (0.026)	0.017 (0.029)	0.0043 (0.026)	0.017 (0.029)
HL high RA estimate									0.12 (0.056)**	0.078 (0.043)*
Controls		x		x		x		x		x
Observations	432	419	431	418	431	418	432	419	432	419

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 45: Predicting smoking status using RA measures (IP7)

	Dependent variable: Smoker ==1 if smokes cigarettes					
HL low RA estimate	-0.047	-0.040				
	(0.048)	(0.048)				
HL high RA estimate			-0.017	-0.015		
			(0.075)	(0.085)		
EG RA estimate					-0.058	-0.052
					(0.036)	(0.041)
Controls		x		x		x
Observations	467	451	467	451	467	451

	Dependent variable: Smoker ==1 if smokes cigarettes									
SOEP-G RA	-0.088	-0.069					-0.089	-0.070	-0.090	-0.070
	(0.036)**	(0.041)*					(0.035)**	(0.039)*	(0.034)**	(0.038)*
SOEP-F RA			-0.0076	0.034						
			(0.041)	(0.041)						
SOEP-H RA					-0.075	-0.043				
					(0.026)***	(0.031)				
HL low RA estimate							-0.042	-0.042		
							(0.051)	(0.051)		
EG RA estimate							-0.034	-0.037	-0.036	-0.039
							(0.040)	(0.046)	(0.041)	(0.046)
HL high RA estimate									-0.028	-0.029
									(0.080)	(0.091)
Controls		x		x		x		x		x
Observations	433	420	432	419	432	419	433	420	433	420

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 46: Predicting heavy drinking using RA measures (IP6) - Ordered Probit

	Dependent variable: How many times drank 5 or more drinks					
HL low RA estimate	-0.016 (0.025)	-0.015 (0.025)				
HL high RA estimate			0.0011 (0.048)	0.0025 (0.042)		
EG RA estimate					-0.0043 (0.026)	0.0053 (0.026)
Controls		x		x		x
Observations	486	485	487	486	486	485

	Dependent variable: How many times drank 5 or more drinks									
SOEP-G RA	-0.079 (0.030)**	-0.032 (0.030)					-0.079 (0.030)**	-0.031 (0.030)	-0.080 (0.030)**	-0.033 (0.030)
SOEP-F RA			-0.099 (0.024)***	-0.067 (0.027)**						
SOEP-F RA					-0.12 (0.023)***	-0.093 (0.027)***				
HL low RA estimate							-0.010 (0.026)	-0.013 (0.026)		
EG RA estimate							0.0024 (0.028)	0.0088 (0.027)	-0.00056 (0.028)	0.0062 (0.028)
HL high RA estimate									0.019 (0.047)	0.0091 (0.043)
Controls		x		x		x		x		x
Observations	488	487	487	486	487	486	488	487	488	487

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 47: Predicting heavy drinking using RA measures (IP7) - Ordered Probit

	Dependent variable: How many times drank 5 or more drinks last 4 weeks					
HL low RA estimate	-0.028	-0.028				
	(0.036)	(0.037)				
HL high RA estimate			0.0068	-0.051		
			(0.067)	(0.074)		
EG RA estimate					0.0084	-0.004
					(0.039)	(0.042)
Controls		x		x		x
Observations	344	335	344	335	344	335

	Dependent variable: How many times drank 5 or more drinks last 4 weeks									
SOEP-G RA	-0.092	-0.044					-0.093	-0.044	-0.093	-0.045
	(0.030)***	(0.031)					(0.031)***	(0.032)	(0.030)***	(0.031)
SOEP-F RA			-0.044	-0.0027						
			(0.034)	(0.038)						
SOEP-H RA					-0.11	-0.083				
					(0.025)***	(0.032)**				
HL low RA estimate							-0.031	-0.028		
							(0.036)	(0.037)		
EG RA estimate							0.017	0.0022	0.014	0.0035
							(0.040)	(0.043)	(0.039)	(0.042)
HL high RA estimate									-0.0100	-0.052
									(0.066)	(0.073)
Controls		x		x		x		x		x
Observations	344	335	343	334	343	334	344	335	344	335

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 48: Predicting savings using RA measures (IP6)

	Dependent variable: Log Savings					
HL low RA estimate	-0.19	-0.087				
	(0.088)**	(0.082)				
HL high RA estimate			-0.088	0.060		
			(0.13)	(0.13)		
EG RA estimate					0.072	0.056
					(0.074)	(0.069)
Controls		x		x		x
Observations	244	244	244	244	244	244

	Dependent variable: Log Savings									
SOEP-G RA	0.23	0.17					0.22	0.17	0.24	0.17
	(0.091)**	(0.073)**					(0.092)**	(0.074)**	(0.093)**	(0.070)**
SOEP-F RA			0.22	0.14						
			(0.10)**	(0.082)						
SOEP-H RA					0.20	0.15				
					(0.10)*	(0.075)*				
HL low RA estimate							-0.18	-0.066		
							(0.098)*	(0.088)		
EG RA estimate							0.047	0.022	0.049	0.018
							(0.071)	(0.073)	(0.073)	(0.075)
HL low RA estimate									-0.17	0.022
									(0.15)	(0.14)
Controls		x		x		x		x		x
Observations	231	231	231	231	230	230	231	231	231	231

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 49: Predicting regular savings type using RA measures (IP6) - Ordered probit

	Dependent variable: Regular Savings (categories)					
HL low RA estimate	-0.031	-0.049				
	(0.041)	(0.045)				
HL high RA estimate			0.0012	-0.020		
			(0.074)	(0.074)		
EG RA estimate					-0.0079	0.0044
					(0.038)	(0.040)
Controls		x		x		x
Observations	342	342	342	342	342	342

	Dependent variable: Regular Savings (categories)									
SOEP-G RA	0.012	0.018					0.011	0.018	0.014	0.025
	(0.039)	(0.043)					(0.039)	(0.043)	(0.041)	(0.045)
SOEP-F RA			0.0013	0.016						
			(0.039)	(0.043)						
SOEP-H RA					0.018	0.021				
					(0.041)	(0.042)				
HL low RA estimate							-0.018	-0.040		
							(0.043)	(0.048)		
EG RA estimate							-0.0090	0.0066	-0.010	0.0057
							(0.039)	(0.041)	(0.039)	(0.041)
HL high RA estimate									-0.0009	-0.028
									(0.079)	(0.080)
Controls		x		x		x		x		x
Observations	325	325	325	325	325	325	325	325	325	325

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 50: Predicting savings horizon using RA measures (IP6) - Ordered probit

	Dependent variable: Long term vs short term savings (categories)					
HL low RA estimate	-0.028	-0.019				
	(0.036)	(0.039)				
HL high RA estimate			-0.025	-0.017		
			(0.056)	(0.060)		
EG RA estimate					-0.023	-0.032
					(0.033)	(0.036)
Controls		x		x		x
Observations	342	342	342	342	342	342

	Dependent variable: Long term vs short term savings (categories)									
SOEP-G RA	-0.025	-0.027					-0.027	-0.028	-0.026	-0.029
	(0.035)	(0.036)					(0.036)	(0.038)	(0.036)	(0.038)
SOEP-F RA			-0.0015	-0.0085						
			(0.036)	(0.038)						
SOEP-H RA					0.012	0.011				
					(0.034)	(0.034)				
HL low RA estimate							-0.021	-0.0077		
							(0.038)	(0.042)		
EG RA estimate							-0.029	-0.040	-0.030	-0.042
							(0.035)	(0.038)	(0.035)	(0.037)
HL high RA estimate									-0.014	0.0058
									(0.056)	(0.060)
Controls		x		x		x		x		x
Observations	325	325	325	325	325	325	325	325	325	325

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 51: Predicting having a personal pension using RA measures (IP6)

	Dependent variable: ==1 if has personal pension					
HL low RA estimate	-0.013	-0.012				
	(0.034)	(0.038)				
HL high RA estimate			-0.11	-0.12		
			(0.060)*	(0.068)*		
EG RA estimate					-0.094	-0.10
					(0.027)***	(0.026)***
Controls		x		x		x
Observations	657	655	657	655	657	655

	Dependent variable: ==1 if has personal pension									
SOEP-G RA	-0.0086	0.022					0.0004	0.029	0.0049	0.036
	(0.027)	(0.033)					(0.028)	(0.035)	(0.030)	(0.036)
SOEP-F RA			-0.030	-0.0076						
			(0.029)	(0.033)						
SOEP-H RA					-0.025	-0.020				
					(0.032)	(0.033)				
HL low RA estimate							-0.0018	-0.00041		
							(0.036)	(0.039)		
EG RA estimate							-0.087	-0.10	-0.081	-0.093
							(0.028)***	(0.027)***	(0.027)***	(0.027)***
HL high RA estimate									-0.078	-0.092
									(0.064)	(0.070)
Controls		x		x		x		x		x
Observations	608	606	607	605	607	605	608	606	608	606

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 52: Predicting DOSPERT - Financial using RA measures (IP7)

	Dependent variable: DOSPERT-F score					
HL low RA estimate	-0.51	-0.75				
	(0.47)	(0.41)*				
HL high RA estimate			-0.66	-1.27		
			(0.78)	(0.71)*		
EG RA estimate					-0.29	-0.45
					(0.43)	(0.42)
Controls		x		x		x
Observations	467	451	467	451	467	451

	Dependent variable: DOSPERT-F score									
SOEP-G RA	-0.68	-0.54					-0.66	-0.52	-0.68	-0.54
	(0.13)***	(0.13)***					(0.12)***	(0.13)***	(0.13)***	(0.13)***
SOEP-F RA			-0.80	-0.71						
			(0.17)***	(0.17)***						
SOEP-H RA					-0.41	-0.21				
					(0.18)**	(0.18)				
HL low RA estimate							-0.29	-0.36		
							(0.24)	(0.23)		
EG RA estimate							0.053	0.036	0.037	0.036
							(0.23)	(0.22)	(0.22)	(0.22)
HL high RA estimate									-0.21	-0.49
									(0.33)	(0.29)*
Controls		x		x		x		x		x
Observations	433	420	432	419	432	419	433	420	433	420

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 53: Predicting DOSPERT - Health using RA measures (IP7)

	Dependent variable: DOSPERT-H score					
HL low RA estimate	-0.49	-0.66				
	(0.49)	(0.45)				
HL high RA estimate			-0.47	-1.03		
			(0.82)	(0.78)		
EG RA estimate					-0.23	-0.41
					(0.41)	(0.38)
Controls		x		x		x
Observations	467	451	467	451	467	451

	Dependent variable: DOSPERT-H score									
SOEP-G RA	-0.81	-0.44					-0.79	-0.43	-0.81	-0.44
	(0.17)***	(0.16)***					(0.17)***	(0.16)**	(0.17)***	(0.16)***
SOEP-F RA			-0.68	-0.34						
			(0.15)***	(0.19)*						
SOEP-H RA					-0.85	-0.49				
					(0.18)***	(0.16)***				
HL low RA estimate							-0.26	-0.22		
							(0.24)	(0.22)		
EG RA estimate							0.13	0.073	0.099	0.061
							(0.24)	(0.23)	(0.24)	(0.23)
HL high RA estimate									-0.0077	-0.17
									(0.35)	(0.31)
Controls		x		x		x		x		x
Observations	433	420	432	419	432	419	433	420	433	420

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 54: Cross-validity ordered probit estimations (IP6)

				All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
HL low RA estimate									
- Consistent	0.071	0.057	0.0036						
	(0.029)**	(0.032)*	(0.024)						
- Consistent in HL low				0.062	0.053	-0.0042			
				(0.026)**	(0.032)	(0.023)			
- Consistent or locally consistent HL low							0.057	0.051	-0.00006
							(0.024)**	(0.028)*	(0.021)
Observations	452	451	451	483	482	482	503	502	502
HL high RA estimate									
- Consistent	0.15	0.093	0.00015						
	(0.047)***	(0.050)*	(0.039)						
- Consistent in HL high				0.15	0.085	0.0043			
				(0.044)***	(0.047)*	(0.039)			
- Consistent or locally consistent HL high							0.14	0.087	0.0051
							(0.040)***	(0.044)*	(0.036)
Observations	452	451	451	512	511	511	525	524	524

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 55: Cross-validity ordered probit estimations - EG (IP6)

				All					
	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA	SOEP-G RA	SOEP-F RA	SOEP-H RA
EG RA estimate									
- Consistent	0.053	0.065	0.062						
	(0.023)**	(0.021)***	(0.020)***						
- Consistent in HL low				0.051	0.060	0.065			
				(0.022)**	(0.021)***	(0.019)***			
- Consistent or locally consistent HL low							0.041	0.054	0.056
							(0.021)*	(0.020)***	(0.017)***
Observations	452	451	451	483	482	482	503	502	502
- Consistent				0.042	0.059	0.045			
				(0.023)*	(0.021)***	(0.019)**			
- Consistent or locally consistent HL high							0.038	0.056	0.040
							(0.022)*	(0.022)**	(0.019)**
Observations				512	511	511	525	524	524

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU.

Table 56: Cross-validity estimations - EG RA ranges, locally inconsistent (IP6)

	Interval Regression, EG RA ranges							
	Consistent in HL low				Consistent or locally inconsistent HL low			
HL low RA estimate	0.26 (0.085)***				0.20 (0.075)***			
SOEP-G RA		0.17 (0.076)**				0.13 (0.071)*		
SOEP-F RA			0.21 (0.078)***				0.19 (0.073)**	
SOEP-H RA				0.22 (0.065)***				0.19 (0.058)***
Observations	518	457	456	457	539	485	484	485

	Interval Regression, EG RA ranges - with controls							
	Consistent in HL low				Consistent or locally inconsistent HL low			
HL low RA estimate	0.25 (0.086)***				0.20 (0.077)**			
SOEP-G RA		0.14 (0.082)*				0.12 (0.079)		
SOEP-F RA			0.17 (0.082)**				0.16 (0.077)**	
SOEP-H RA				0.19 (0.068)***				0.17 (0.063)***
Observations	517	456	455	456	538	484	483	484
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 57: Cross-validity estimations - EG RA ranges, locally inconsistent (HL high, IP6)

	Interval Regression, EG RA ranges							
	Consistent in HL high				Consistent or locally inconsistent HL high			
HL high RA estimate	0.40				0.39			
	(0.11)***				(0.11)***			
SOEP-G RA		0.13				0.12		
		(0.077)*				(0.074)		
SOEP-F RA			0.19				0.18	
			(0.075)**				(0.077)**	
SOEP-H RA				0.15				0.13
				(0.064)**				(0.062)**
Observations	546	492	491	492	562	507	506	507

	Interval Regression, EG RA ranges - with controls							
	Consistent in HL high				Consistent or locally inconsistent HL high			
HL high RA estimate	0.43				0.42			
	(0.11)***				(0.11)***			
SOEP-G RA		0.11				0.091		
		(0.079)				(0.077)		
SOEP-F RA			0.17				0.15	
			(0.078)**				(0.080)*	
SOEP-H RA				0.12				0.10
				(0.069)*				(0.067)
Observations	545	491	490	491	561	506	505	506
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 58: Cross-validity estimations - HL low RA ranges, locally inconsistent (IP6)

	Interval Regression, HL low RA ranges							
	Consistent in HL low				Consistent or locally inconsistent in HL low			
EG RA estimate	0.15				0.14			
	(0.047)***				(0.049)***			
SOEP-G RA		0.12				0.13		
		(0.056)**				(0.059)**		
SOEP-F RA			0.10				0.12	
			(0.069)				(0.070)*	
SOEP-H RA				0.0023				0.011
				(0.049)				(0.054)
Observations	518	457	456	457	539	485	484	485

	Interval Regression, HL low RA ranges - with controls							
	Consistent in HL low				Consistent or locally inconsistent in HL low			
EG RA estimate	0.14				0.13			
	(0.048)***				(0.050)***			
SOEP-G RA		0.13				0.15		
		(0.067)*				(0.068)**		
SOEP-F RA			0.10				0.13	
			(0.074)				(0.077)	
SOEP-H RA				0.00068				0.014
				(0.051)				(0.057)
Observations	517	456	455	456	538	484	483	484
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

Table 59: Cross-validity estimations - HL high RA ranges, locally inconsistent (IP6)

	Interval Regression, HL high RA ranges							
	Consistent in HL high				Consistent or locally inconsistent in HL high			
EG RA estimate	0.11 (0.031)***				0.13 (0.048)***			
SOEP-G RA		0.12 (0.039)***				0.11 (0.059)*		
SOEP-F RA			0.082 (0.044)*				0.079 (0.069)	
SOEP-H RA				0.014 (0.035)				0.024 (0.051)
Observations	546	492	491	492	562	507	506	507

	Interval Regression, HL low RA ranges - with controls							
	Consistent				Consistent or locally inconsistent in HL high			
EG RA estimate	0.11 (0.032)***				0.12 (0.049)**			
SOEP-G RA		0.15 (0.043)***				0.13 (0.069)*		
SOEP-F RA			0.097 (0.045)**				0.079 (0.077)	
SOEP-H RA				0.028 (0.035)				0.031 (0.058)
Observations	545	491	490	491	561	506	505	506
Controls	x	x	x	x	x	x	x	x

Notes - Standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions use sampling weights. Standard errors account for strata and PSU. Controls include: Gender, Age, Marital Status, Education, Income (ln), Employment Status and Interview mode.

B Innovation Panel Wave 6 field procedures and treatments

The sampling of the IP is as follows. Within each of the eleven GORs (with Wales and Scotland each counted as a single region), sectors were sorted in order of the percentage of "household heads" (defined as the "household reference persons" responsible for owning or renting, or who are otherwise responsible for the accommodation) classified as National Statistics Socio-Economic Classification (NS-SEC) categories 1 ("Higher managerial, administrative and professional occupations") and 2 ("Lower managerial, administrative and professional occupations"), and then divided into three approximately equal-sized bands, thus creating 33 strata. Within each of those strata, sectors were then sorted in order of population density. Treating the list of sectors as a single ordered list, a sample of 120 sectors was selected systematically with probability proportional to address count. Using a fixed interval and a random start, 23 addresses were selected systematically from each sampled sector, making a total sample of 2,760 addresses from 120 sectors. The IP sample design is thus described by the sampling stratum, the PSUs, and the design weights. In the estimations we make use of the survey weights to adjust for differential nonresponse and to generalize results to the UK population (see Section 3).

The final stage of sampling was carried out in the field by the interviewers. For each sampled address, the interviewer identified the sample persons. All persons resident in the sample address at the time the interviewer collected the household grid information for Wave 1 were defined as sample members. A refinement of this procedure was applied at the small number of addresses that contained multiple dwellings or households, primarily to constrain interviewer workloads: if more than three dwellings were found at an address, or more than three households were found to be resident at a sampled dwelling, three were selected at random for inclusion using a Kish grid procedure.

All household members aged 16 years or older were eligible for interview. As with the main *Understanding Society* survey, all persons resident at the address in Wave 1, including children, were defined as "original sample members" (OSM) to be followed throughout the life of the study. In IP Waves 4 and 7, additional 960 and 1,560 new addresses, respectively, were added as "refreshment samples" (8 and 13 in each of the 120 PSUs, respectively).

The IP5 sample had two components: the original sample members (OSMs) that were at their fifth wave of data collection, and the members of the Wave 4 refreshment sample, at their second wave. In Wave 5, households in both samples (n=1,575) were randomly assigned within PSUs to one of two treatment groups: face-to-face (F2F) interview, (n=532, about a third of the sample, of which 168 from the refreshment sample), and web-mode (WEB) interview, (n=1,043, about two-thirds of the sample, of which 315 from the refreshment sample). Because randomization was implemented across PSUs, each sampling point contained a mix of households in each treatment group. Randomization was at a household level: all adult members (aged 16 or above) within the household received the same experimental treatment and any split-off households retained the same experimental allocation from the previous household. As a result, in Wave 5 a total of n=1,293 IP respondents were randomly allocated to the WEB group (n=856 OSMs and n=437 from the refreshment sample) and n=702 to the F2F group (n=459 OSMs and n=243 from the refreshment sample).

Wave 6 replicated this random allocation with households receiving the same experimental treatment as in IP5. As a result, in IP6 a total of n=1,337 IP respondents were randomly allocated to the WEB group (n=910 OSMs and n=427 from the refreshment sample) and n=686 to the F2F group (n=446 OSMs and n=240 from the refreshment sample).

The same original random allocation of households to WEB versus F2F treatments was also carried over in Wave 7. In IP7 a total of n=1,126 IP respondents were randomly allocated to the WEB group (n=773 OSMs and n=353 from the refreshment sample) and n=554 to the F2F group (n=364 OSMs and n=190 from the refreshment sample).

Waves 6 and 7 differed from IP5, however, in that it incorporated a "mop-up" of uncompleted households at the end of the standard fieldwork (see more below).

The F2F group involved standard *Understanding Society* procedures: each adult member was sent an advance letter with the same unconditional incentive (£10), after which interviewers called to attempt computer-assisted personal interviewing (CAPI) interviews.

In each household one adult member was asked to complete the *household grid*, which collects information on who is currently living in the household. Typically the household grid has a question to identify who is responsible for paying utility bills in the household.

This person or their spouse/partner could then complete the *household questionnaire*. The household questionnaire is relatively short (around 10 minutes) and collects household-level information such as housing tenure, rent/mortgage payments, expenditures, utility bills, household consumer durables, and some measures of material deprivation. All household adult members were then asked for an individual interview and to complete a self-completion questionnaire, which was randomly allocated to be either a computer-assisted self-interview (CASI) or a paper questionnaire booklet. Young people aged 10-15 were administered a paper self-completion questionnaire.

Fieldwork procedures for the WEB treatment were as follows: all adult members were sent an advance letter with a financial incentive, inviting them to take part to the web interview. The value of the incentives was also subject to some experimental random allocation. Households in the WEB group were randomly allocated in equal proportions to either an unconditional £10 incentive (WEB10); or an unconditional £30 incentive (WEB30); or, finally, an unconditional £10 incentive, plus extra £20 conditional on full-household completion by web in a two-weeks "web-only" allotted time (WEB10+20).

The advance letter included the URL and a unique user ID which was to be entered on the welcome online screen. A version of the letter was additionally sent by email to all adult members for whom an email address was available. For members who, in previous waves, had indicated that they did not use internet regularly for personal use, the letter said they would also have the opportunity to do the survey with an interviewer. Up to the three email reminders were sent at 3-day intervals. Adult members who had not completed the web interview after the two weeks were sent a reminder by post and interviewers started visiting them to carry out CAPI interviews. This was the two-week "web-only" period in which households allocated to the WEB10+20 group could qualify for the additional £20 incentive.

The first household member to log on to complete the web survey was asked to complete the household grid. In the web mode the household grid also had a question to identify who was responsible for paying bills in the household. This person or their spouse/partner could then complete the online household questionnaire.

In general, the adult web questionnaire was substantially based on the CAPI version with some appropriate adaptations, such as incorporating interviewer instructions into question wording, and making "help" screens more respondent-friendly. The web survey was not suitable for completion using a smart phone or any other small mobile device. If a smart phone was used to access the log-on page, the respondent was automatically directed to a page requesting that they log on from a computer.

The WEB survey remained open throughout all the standard F2F fieldwork period. Interviewers informed household members in the WEB group that the interview would be closed approximately one week before the end of the fieldwork. In Waves 6 and 7 (but not Wave 5) households who had not been surveyed at the end of the standard fieldwork period, and were not adamant refusals, were contacted again in a "mop-up" stage of fieldwork.

The type of "mop-up" contact was different across groups. The F2F group was contacted to offer a web interview during the mop-up stage. All individuals were sent a letter including the URL of the web and their unique user ID which was to be entered on the welcome online screen. A version of the letter was additionally sent by email to the adult members for whom an email address was available. Several days later, a telephone interviewer contacted all those respondents for whom phone number was known in order to remind them of the web questionnaire, and to administer a telephone interview if possible.

In the WEB group, the "mop-up" contact was made by telephone. The telephone interviewer reminded the household member that they could participate on the web, but that was also able to administer the interview by telephone (CATI). Cases for which a telephone number was not known were not contacted again at the mop-up stage.

C Allocation to the risk preferences module

The initial random allocation of households eligible to the risk preferences experimental module across the two interview modes changed because of the above described two-weeks "web-only" period and the related "mop-up" procedure. In particular, during that period, 228 of the 528 households originally allocated to the WEB mode (43.2% of the original WEB eligible households) opted for the F2F mode. In the "mop-up" session, further 11 of the 528 households originally allocated to the WEB mode (2.08% of the original WEB eligible households) had a telephone interview. Because the experimental module was impossible to administer in the CATI mode, these 11 subjects therefore lost eligibility for the experimental module. The same happened to 3 of the 280 households initially allocated to the F2F mode (1.07% of the original F2F eligible households), who also had a CATI interview in the "mop-up" session, and therefore lost eligibility for the experimental module. There were also 6 of the 280 households initially allocated to the F2F mode (2.14% of the original F2F eligible households) who in the "mop-up" session ended up completing a web interview instead.

As a result, at the end of the two-week "web-only" period and of the "mop-up" sessions, there were n=794 households potentially eligible for the experimental module (the original 808, minus the 14 interviewed in the CATI mode): n=499 of them were in the F2F interview mode (62.85% of the total), and n=295 in the WEB mode (37.15% of the total).

In each household eligible for the experimental module, one adult respondent in the household was then randomly selected with a Kish grid of enumerated adults. The random selection was done to ensure a fair representation of both spouses and, more generally, of all adults in the household. Only one adult respondent thus answered the risk preferences questions in each household.

D Preamble to risk preferences module

To guarantee clarity and transparency of the procedure, and to ensure that subjects understood the consequences of their choices, the risk (and time) preferences module started with the following instructions preamble:

> "We will ask you a set of very brief questions. At the end you will have the chance to win an amount of money. In each question, you will be presented with a choice. There are two sets of questions. [$\cdots$] In the second set questions you will be asked to decide between types of lottery games that you might prefer to play. In all these questions, there is no right or wrong answer. We are interested in what you genuinely prefer."

At the end of this preamble, respondents were also explicitly asked whether they intended to continue with the experimental task, or whether they preferred to refuse to take part in it. A similar outside option was presented before the risk preferences questions (e.g. some respondents may not be willing to play lottery gambles for personal or religious reasons).

To guarantee that subjects understood the experimental procedure and were fully aware of the consequences of their choices, the above instructions preamble further explained:

> "Once you have made all your choices, you could receive a payment. Ten percent of respondents will be randomly selected to win. There are 91 questions in this section of the interview in total. If you are selected to receive the payment, one of the 91 questions will be selected at random, and your payment will be calculated based on your answer to that question. The highest amount that you could win is £250. If you are selected to receive the payment, the payment will be arranged at the end of the interview."

E Randomized payments administration

Randomized payments were administered by mean of an *ad hoc* computerized script that was implemented in both F2F and WEB response modes. In the script, a screen of instructions reminded subjects that they had "1 in 10 chance to receive a payment." It further explained that "If you are selected to receive the payment, one of the 91 questions that you just answered will be selected at random, and the payment will be calculated based on your answer to that question."

A screen shot then simulated a roll of a ten-faced die to see whether the respondent was actually selected to receive the monetary payment for real. If not, a screen with a "Sorry, you have not been selected to receive the payment" message appeared, and the participant moved on to the final questionnaire, described above. If so, a screen with a "Congratulation, you have been selected to receive the payment" appeared with the further explanation that "Let's now see which question will be selected to calculate the payment. A number between 1 and 91 will be selected at random to choose the question for your payment." Another screen shot followed that simulated a random draw of a ball from an urn containing 91 numbered balls, one for each risk (and time) preferences task.

The sequence of the next screens then depended on the type of experimental questions that was randomly extracted. If one of the binary lottery questions was randomly selected (balls numbered 73 to 90), a screen appeared explaining that the question corresponding to that number was selected, and reporting the whole wording of the selected question. The screen also reported which lottery was actually chosen by the respondent. A subsequent screen simulated a roll of a ten-faced die to see which outcome was selected within the preferred lottery. Another subsequent screen then summarized what was the outcome of the preferred lottery and that, at the end of the interview, the professional interviewer would give the respondent a voucher for the corresponding amount. The procedure was analogous in case the B-EG risk preferences test was selected (ball numbered 91), with the only difference that the summary screen after the die-roll summarized whether the *low* or *high* outcome had occurred with the corresponding payment.

F SOEP and DOSPERT Questions

F.1 Socio-Economic Panel (SOEP)

The Socio-Economic Panel questions regarding risk attitudes are presented in separate frames, as shown in Figures 4-6.

Figure 4: SOEP General Risk Question

Are you generally a person who is willing to take risks or do you try to avoid taking risks?

Please answer on a scale from 0 to 10, where
0 means "I am generally a person unwilling to take risks" and
10 means "I am generally a person fully prepared to take risks"

Unwilling to take risks										Fully prepared to take risks
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10

Figure 5: SOEP Financial Risk Question

Are you generally a person who is willing to take risks or do you try to avoid taking risks in financial matters?

Please answer on a scale from 0 to 10, where
0 means “I am generally a person unwilling to take risks in financial matters” and
10 means “I am generally a person fully prepared to take risks in financial matters”

Unwilling to take risks										Fully prepared to take risks
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10

Figure 6: SOEP Health Risk Question

Are you generally a person who is willing to take risks or do you try to avoid taking risks in health?

Please answer on a scale from 0 to 10, where
0 means “I am generally a person unwilling to take risks in health” and
10 means “I am generally a person fully prepared to take risks in health”

Unwilling to take risks

Fully prepared to take risks

□	□	□	□	□	□	□	□	□	□	□
0	1	2	3	4	5	6	7	8	9	10

F.2 Domain-Specific Risk-Taking Scale (DOSPERT)

The Domain-Specific Risk-Taking Scale questions are presented in separate frames. The first question is shown in Figure 7. The full set of questions is presented in Table 60.

Figure 7: DOSPERT Question 1

For each of the following statements, please indicate the likelihood that you would engage in the described activity or behaviour if you were to find yourself in that situation. Provide a rating using the following scale, where 1 is “Extremely Unlikely” and 7 is “Extremely Likely”. Do not spend too much time on any statement. Please answer quickly and honestly.

Betting a day’s income at the horse races

1	Extremely unlikely
2	Moderately unlikely
3	Somewhat unlikely
4	Not sure
5	Somewhat likely
6	Moderately likely
7	Extremely likely

Table 60: Domain-Specific Risk-Taking Scale (DOSPERT) Questions

1	Betting a day's income at the horse races
2	Investing 10% of your annual income in a moderate growth mutual fund
3	Betting a day's income at a high-stake poker game
4	Investing 5% of your annual income in a very speculative stock
5	Betting a day's income on the outcome of a sporting event
6	Engaging in unprotected sex
7	Driving a car without wearing a seatbelt
8	Investing 10% of your annual income in a new business venture
9	Riding a motorcycle without a helmet
10	Sunbathing without sunscreen
11	Walking home alone at night in an unsafe area of the town

G Responses in IP7 and across IP6 and IP7

Responses in IP7

Also in IP7 our sample of respondents provided heterogeneous responses in the B-EG task. More than a third of the 468 subjects who answered the B-EG task in IP7 (n=167, 35.68% of the respondents) chose certain lottery A in B-EG. The riskiest lottery F was chosen only by 54 subjects, 11.54% of the total. A total of 101 respondents (21.58%) chose lottery C.

Also in IP7 there is wide heterogeneity in the responses to the HL task as well. First, a total of 188 subjects out of 468 (40.17% of the IP7 sample, compared to 31.01% in IP6) made at least one inconsistent choice in the sense of switching back to lottery A after having chosen lottery B (32.47% and 23.5% in the low and high stakes questions). There is a core group of 74 respondents, who made at least one inconsistent choice in both the low and the high monetary stakes HL questions.

Less than a fifth of the respondents never chose the safer lottery A: 17.09% with low monetary stakes and 18.38% with high stakes. A roughly equal proportion of subjects always chose the safer lottery A in all series of choices (16.03% and 21.58% with low and high stakes, respectively).

Responses across IP6 and IP7

For 413 respondents we are able to link data across the IP6 and IP7 waves. This subsample of IP6 subjects provided responses in the IP6 B-EG task that were partly different from the responses of the overall IP6 sample. In particular, while more than a third of those 413 subjects (36.32% of the respondents) chose the certain lottery A, in B-EG (compared to 35.11% of the whole IP6 sample), only 10.99% of them (compared to 15.88% of the whole IP6 sample) chose the riskiest lottery F. A total of 91 from those respondents (22.03%) chose lottery C (compared to 18.32% of the whole IP6 sample).

A total of 125 of those subjects (30.43% of the total) made at least one inconsistent choice when answering HL questions in IP6 (in the sense that they switched back), in line with the analogous proportion in the whole IP6 sample (31.01%). In particular, when answering HL questions with low monetary stakes, 23.49% of the sample made inconsistent choices, while 18.16% behaved as switchers with high monetary stakes. There is a core group of 48 respondents, who made at least one inconsistent choice in both the low and the high monetary stakes HL questions.

Less than a fifth of the respondents who answered HL questions in both IP6 and IP7 never chose the safer lottery A in IP (16.95% with low monetary stakes, 19.13% with high stakes). A roughly equal proportion of subjects always chose the safer lottery A in all series of choices in IP6 (15.98% and 20.82% with low and high stakes, respectively). The overall distribution of IP6 responses to the HL questions for the subjects who answered the HL questions in both IP6 and IP7 is thus fairly comparable with the corresponding distribution for the whole IP6 sample, except for the smaller proportions on the polar cases of no or all lottery A choices.