Experts in Experiments:
How Selection Matters for Estimated Distributions of Risk Preferences

Hans-Martin von Gaudecker, Arthur van Soest, and Erik Wengström[*]

March 12, 2011

Abstract

An ever increasing number of experiments attempts to elicit risk preferences of a population of interest with the aim of calibrating parameters used in economic models. We are concerned with two types of selection effects, which may affect the external validity of standard experiments: Sampling from a narrowly defined population of students ("experimenter-induced selection") and self-selection of participants into the experiment. We find that both types of selection lead to a sample of experts: Participants perform significantly better than the general population, in the sense of fewer violations of revealed preference conditions. Self-selection within a broad population does not seem to matter for average preferences. In contrast, sampling from a student population leads to lower estimates of average risk aversion and loss aversion parameters. Furthermore, it dramatically reduces the amount of heterogeneity in all parameters.

---

1 Introduction

Preference heterogeneity is pervasive and crucially changes the implications of many economic models (Browning *et al.*, 1999; Blundell and Stoker, 2007). Estimating its nature from observational data often requires questionable assumptions on market completeness, information structure, etc. As a consequence, there is increasing interest in studying individual heterogeneity using tools from experimental economics.[1] The tightly controlled setting of experiments makes revealed preference arguments more credible and the estimated distributions of parameters can be used to calibrate the parameters of economic models. A major threat to the estimates' external validity arises, however, if the population from which the participants in the experiment are drawn is not the same as the population of interest.[2] What constitutes the population of interest may vary between studies, but in many cases researchers ultimately seek to make statements about human behaviour in general.[3] As a result, it is important to assess how well the behaviour of experimental subjects corresponds to the behaviour of the population at large.

In this paper, we exploit two prominent sources of differences between the sample of subjects in an experiment and the broad population of interest. First, the standard recruitment protocol in experimental economics – researchers inviting college students via emails or posters – may restrict socio-demographic variation too severely (Harrison *et al.*, 2002). We coin this effect experimenter-induced selection. Second, even if those who are invited to participate form a random sample of the population of interest, voluntary participation potentially gives rise to systematic self-selection (Heckman, 1974).

Recent years have witnessed different approaches to enhance demographic variation in experimental situations. A popular method is to take the laboratory to the population of interest; for prominent examples along these lines see the series of papers Harrison and coauthors (Harrison *et al.*, 2002, 2007a,b; Andersen *et al.*, 2008, 2010). Experimenter-induced selection is minimised, but self-selection remains a concern (Harrison *et al.*, 2009). The relatively high costs of administering experiments in this fashion have led to moderate sample sizes. Another strategy to minimise experimenter induced selection that has become available recently is to integrate experiments into (pilots of) existing household surveys; see for example the pioneering work by Fehr *et al.* (2003) or Dohmen *et al.* (2005, forthcoming). A further advantage of this is the availability of a large amount of background information on participants, which helps in estimating the effects of self-selection. Until now, capacity constraints in the survey instruments have prevented a more widespread use of this method. A third approach has employed large convenience samples of Internet respondents recruited by means of newspaper advertising or email invitations. Examples include Lucking-Reiley (1999) and

---

[1] A non-exhaustive list of recent contributions includes Harrison *et al.* (2002); Bleichrodt *et al.* (2001); Dohmen *et al.* (forthcoming); Choi *et al.* (2007); Andersen *et al.* (2008); Tanaka *et al.* (2010); von Gaudecker *et al.* (2011).

[2] Other threats to external validity may occur if the experimental design differs too widely from real-world situations in terms of context, stakes, or similar features. See Harrison and List (2004) and Levitt and List (2007) for reviews of such effects and Falk and Heckman (2009) for critical thoughts.

[3] While this is true for most studies, there are clearly arguments for being interested in preference estimates from specific samples of the population. For example, Lazear *et al.* (2011) have recently argued that in certain settings self-selection on preferences can create market outcomes that are very different from those that would have occurred if the general population participated. In order to understand or predict behaviour in such situations, one would naturally rather have preference estimates of those active on the specific markets rather than the population at large.

Güth *et al.* (2007). Experimenter-induced selection is a major concern here since respondents have to read a particular newspaper, need to be able to access the Internet, have to subscribe to a specific electronic mailing list, etc. Since there is no information on non-participants, accounting for self-selection effects is virtually impossible.

We combine features of these three approaches in analysing an experiment with a large sample of respondents from a Dutch household survey, the CentERpanel. This is an Internet survey with a sample drawn from a population register, which avoids non-coverage of those without Internet access by providing them with the necessary equipment to participate. In order to investigate the importance of experimenter-induced selection, we compare the Internet outcomes to those of parallel laboratory experiments. As we do not observe the CentERpanel subjects in a laboratory, experimenter-induced selection effects might be confounded with implementation mode effects. We address this issue from two angles. First, we introduce a treatment in the laboratory which replicates the Internet setting as closely as possible. Second, our Internet sample is sufficiently large to analyse a subsample that resembles the student population in terms of age and education. If environmental factors play a role, they should lead to differences between the results for this Internet subsample and those for the laboratory experiment.

Although the CentERpanel is a random sample *ex ante*, the fact that not everyone who is invited to participate indeed completes the experiment may still lead to (self-)selection bias. In order to quantify this, we exploit the rich amount of background information from the existing survey, which is available for participants as well as non-participants and analyse self-selection under a missing at random assumption (MAR; see Little and Rubin, 2002).

Our empirical analysis is structural in the sense that we specify a model of optimal behaviour in the choice tasks faced by participants and estimate its primitive parameters. We bring the model to the data by means of a random coefficients model, which naturally allows for observed and unobserved heterogeneity in all preference parameters and in the tendency to make optimisation errors. The latter are identified because we have a large number of choices available for each individual. Earlier research has pointed out that error rates can be substantial and that the propensity to make errors varies with socio-economic background variables (von Gaudecker *et al.*, 2011). Our preferred version of the behavioural model contains parameters for utility curvature and loss aversion (Kahneman and Tversky, 1979), but we exploit several alternatives in the Online Appendix, demonstrating that our conclusions are robust to the specific form of risk preferences assumed.

Our findings indicate that there are no effects of implementation mode. These results for risk preferences are largely in line with what Bellemare and Kröger (2007) find for trust: students are not representative for the broad population, but once individual characteristics are controlled for, behaviour in a trust game in the lab and over the Internet are not significantly different. Student samples lead to smaller average parameter estimates for risk aversion and loss aversion. Error rates and heterogeneity in all parameters are dramatically reduced compared to the general population. This suggests that it is not possible to draw inference on risky choice behaviour and preference heterogeneity in the broad population from narrowly defined student samples. These effects seem to be driven entirely by the choice of the sampling population rather than the implementation mode. Compared to the effects of experimenter-induced selection on average parameter estimates, we find effects of self-selection that are small to non-existent. We find an economically meaningful effect only for the error propensity. Hence, it seems that ignoring self-selection does far less harm than sampling from

a narrowly defined distribution.

Our results have important implications for measuring risk preferences. First, parameter distributions estimated on the basis of a classical experiment using a self-selected student sample cannot simply be extrapolated to a broader population of interest. Second, an Internet experiment using a sound sampling frame based upon a probability sample from the population of interest and also covering the population without Internet access can be an excellent alternative. While such an experiment in principle might also suffer from selection effects (due to voluntary participation) as well as implementation mode effects (lab versus Internet), our results demonstrate that these biases are negligible compared to the biases due to differences between a students only and a population representative sample.

2 Experimental setup and theoretical specification

This section provides a detailed description of our experimental design, the sampling populations and our econometric model. The starting point of the experiments is the multiple price list format, a well-established methodology for preference elicitation. We first describe the multiple price list format and how we implement it. We then point out the aspects of the experiment that are specific to the Internet and laboratory settings, respectively. In particular, we highlight the features of our design aimed at disentangling the effects of experimenter-induced selection and implementation method. One of these features is the introduction of two environmental treatments in the laboratory. The first treatment replicates traditional experiments, the other one mimics the Internet setting as much as possible. We term them "Lab-Lab" and "Lab-Internet" to avoid confusion with the CentERpanel experiment (also denoted as "Internet experiment"). The full experimental instructions, samples of choice problems, help screens, final questions, and the debriefing part are available in the Online Appendix. The Internet experiment is also used in von Gaudecker *et al.* (2011). Section 2.4 sketches the behavioural model and its empirical implementation.

2.1 The Multiple Price List format

Our experiments use a modified version[4] of the multiple price list format; see, e.g., Binswanger (1980), Holt and Laury (2002) and Andersen *et al.* (2006). Subjects get series of lotteries with identical payoffs but varying probabilities such as those in the screenshot in Figure 1, taken from (von Gaudecker *et al.*, 2011). In each of the four cases, the subject chooses between Options 'A' and 'B'. The lotteries are designed such that according to all economic theories that we consider (see below), subjects who do not make mistakes switch at some point from the safer Option 'A' to the riskier Option 'B', or choose 'B' throughout (since the last row is always a choice between certain payoffs, higher for 'B' than for 'A'). Probabilities of the high payoff on the first screen increase from 0.25 to 1 in steps of 0.25. To obtain more precise information on switching points, subjects who are consistent in the sense that they do not switch back and forth between 'A' and 'B' and choose the higher certain payoff in the final question are routed to a second screen, containing lotteries with the same payoffs but a refined probability grid – involving 10%-steps located approximately between the respondent's highest choice of

---

[4]The main modification is that we include pie charts as a graphical illustration of probabilities (cf. Figure 1); see (von Gaudecker *et al.*, 2011) for details.

Figure 1: Screenshot of payoff configuration 5, first screen

Table 1: Characteristics of the Seven Payoff Configurations

Payoff Configuration	Option A			Option B		
	Uncertainty Resolution	Low Payoff	High Payoff	Uncertainty Resolution	Low Payoff	High Payoff
1	early	27	33	early	0	69
2	early	39	48	early	9	87
3	early	12	15	early	-15	48
4	early	33	36	late	6	69
5	early	18	21	late	-9	54
6	early	24	27	early	-3	60
7	late	15	18	late	-12	51

Note: Subjects were asked to choose between the two options on each row 4 or 8 times, with different probabilities of obtaining the high outcome. These values were shown in the high incentive and hypothetical treatments. For the low incentive treatment they were divided by three. The order was randomised.

'A' and lowest choice of 'B' on the previous screen, similarly as in the iterative multiple price list format described in Andersen *et al.* (2006). For example, if the subject switches between 0.25 and 0.50, the second screen has probabilities 0.2, 0.3, 0.4 and 0.5.

Each subject faced the seven payoff configurations described in Table 1;[5] for each configuration they make four or eight decisions, depending on whether they get the second screen. While 'A' always guarantees a positive payoff, Some of the 'B' lotteries involved a possible loss. Actual payments were always made three months after the experiment. At the top of each screen, we indicated whether the outcome of the lottery was revealed immediately or in three months' time (see Figure 1).

Subjects were randomly split into three groups: groups with hypothetical and real lotteries with the amounts shown in Table 1, and a third group with real payoffs but amounts divided by three. We refer to these groups as hypothetical, high, and low incentive treatments. Subjects in the high and low incentive groups received an upfront payment of 15 or 5 Euros, respectively, if they completed the experiment. These amounts were chosen to compensate for the maximum negative payoffs in the lotteries. No payment was made to the hypothetical group of the CentERpanel experiment, but to recruit the laboratory subjects in the hypothetical group, these were given a participation fee of 5 Euros. In the incentives treatments, everyone received the participation fee, but only one in ten subjects additionally got paid for one of the chosen lotteries (as in Harrison *et al.* (2007a) and Dohmen *et al.* (forthcoming)). The lottery to be paid out was selected at random to ensure incentive compatibility. We randomised the order in which the seven payoff configurations were presented.

---

[5]The payoff configurations are scaled and rounded versions of those used by Holt and Laury (2002).

2.2 The Internet experiment

The subjects in the Internet experiment are respondents in the CentERpanel,[6] aged 16 and older. The CentERpanel is managed by CentERdata, a survey research institute affiliated with Tilburg University. The panel contains more than 2,000 households and covers the complete Dutch population, excluding the institutionalised. Questionnaires and experiments are fielded over the Internet. To avoid selection bias, households without Internet access are provided with a set-top box for their TV (and a TV if necessary). Panel members get questions every weekend. They are reimbursed for their costs of participation (fees of dial-up Internet connections etc.) on a regular basis. We conducted our experiments in November and December of 2005 and payments were made with the regular transfers three months after the experiments. The data for the Internet experiment were also used in von Gaudecker *et al.* (2011); see that paper and the online appendix for more details, screenshots, etc.

The welcome screen contained a brief introduction to the experiment followed by a non-participation option (see Figure 6 in the Online Appendix for the introductory screens of all treatments). For the treatments with real incentives, subjects were told the amount of the participation fee and that they had the chance to win substantially more or lose (part of) this money again. It was made clear that no payment would be made upon non-participation or dropping out. In the hypothetical treatment, subjects were informed that the questionnaire consisted of choices under uncertainty in a hypothetical setting. In all treatments, subjects then had to indicate whether they wanted to participate or not. Respondents who opted for participation first went through two pages of online instructions before facing the seven price list configurations. The instructions and specially designed help screens could be accessed throughout the experiment. They were included to improve comparability with similar laboratory experiments, compensating for the absence of an experimenter.

In total, 2,299 persons logged into the system. About 12.7% opted for non-participation, leaving 2,008 respondents who started the experiment. 80 subjects dropped out before completing the questionnaire. Moreover, 138 respondents went through the experiment extremely rapidly. Those who took less than 5:20 minutes (the minimum completion time observed in the laboratory experiments) are treated as dropouts in the analysis below (see also Section 4). Our final sample thus consists of 1,790 subjects who made 91,808 choices.

The first three columns of Table 2 list descriptive statistics for the participants who completed the experiment ("final sample"), those who opted for non-participation, and those who dropped out in the course of the experiment or sped through it. As expected, the three groups differ in many respects. The variables in Table 2 can be broadly classified into six groups: Incentive treatment; education; sex and age; employment status and residential area; financial literacy and experience with financial decision making; income. Not everybody answered all of the questions, implying smaller sample sizes if we include all variables. Item nonresponse is strongest for the questions on assets and financial literacy and experience.

Respondents randomised into the hypothetical treatment are underrepresented in the final Internet sample; they more often decide not to participate and more often drop out or speed through. Those who got the high incentive treatment are particularly unlikely to drop out or

---

[6]For related papers using data collected through the CentERpanel see, e.g., Donkers *et al.* (2001) who analysed risk preferences using hypothetical questions, and Bellemare and Kröger (2007) and Bellemare *et al.* (2008) for evidence from a trust game and an ultimatum game with real payoffs. For more information about the CentERpanel, see `http://www.centerdata.nl/en`.

speed through the Internet experiment, as one would expect. Non-participation is negatively related to education level and positively associated with age. On the other hand, dropping out or speeding through is more likely for the younger age groups. Women are less likely to complete the Internet experiment than men; they more often decide not to participate and also more often drop out or speed through.

Respondents with high household income participate more often. Participation also rises with several indexes of experience with financial decision making, such as being the person mainly responsible for household finances or having stocks. We also included two variables referring to employer provided savings plans. This is a specific type of long term savings plan that is heavily subsidised by the government through tax deductions, see Alessie *et al.* (2006), making net returns much higher than on any other safe asset. While it is easy to sign up for these plans and the employer does most of the paperwork, the default is not to participate. This may explain why employees with little financial knowledge or interest often do not sign up; see, for example, the work of Madrian and Shea (2001) on non-take up of 401(k) plans. Conditional on being offered such a plan, participating in it can be seen as an index of financial literacy. The fact that non-participation and dropping out or speeding through are negatively associated with such a plan therefore suggests that completing the experiment is positively associated with financial literacy.

2.3 The Laboratory Experiment

In order to compare the answers in the Internet experiment to those from a "standard" laboratory experiment, we performed the same experiment in the economics laboratory at Tilburg University. In total, 178 students participated in 16 sessions, divided equally between September 2005 and May 2006. The same treatments were carried out as in the Internet survey. The only difference was the payment of a show-up fee in the hypothetical treatment mentioned in Section 2.1. The payment procedure for the incentives treatments was as in the CentERpanel experiment: The participation fee was transferred to participants' bank accounts three months after the experiment; one in ten subjects received the sum of the participation fee and the (possibly negative) payment from one randomly drawn lottery.

To distinguish effects due to different sampling populations from effects due to replacing the controlled laboratory setting by the Internet environment, we also replicated this latter change in the lab. The first environmental treatment, labelled the "Lab-Lab" treatment, replicates the traditional setup used in laboratory experiments. In particular, an experimenter was present in the room to help the subjects and answer questions. In contrast to the CentERpanel experiment, no links to the instructions or help screens were shown in the core part of the experiment. Otherwise, the screens resembled the one in Figure 1. Participants also had to wait until everyone else in the session had finished before they could leave the room. In the second environmental treatment – termed the "Lab-Internet" treatment – the experimenter was not present. Instead subjects had access to the same help screens (including the introductory screens) as in the CentERpanel experiment. Moreover, subjects could leave directly after completing the experiment – they did not have to wait for everyone else.

The last column of Table 2 contains the available demographic characteristics of the laboratory subjects. Compared to the CentERpanel experiment, there is less information and less variation in the basic demographic characteristics. Specifically, in terms of age and education, the laboratory subjects cover only a small fraction of the population represented

Table 2: Selected Characteristics of Participants

Variable	Final Sample	CentERpanel Non-Participants	Dropouts Speeders	Laboratory Final Sample
Hypothetical treatment	0.31	0.50	0.37	0.37
Low incentive treatment	0.37	0.23	0.35	0.27
High incentive treatment	0.32	0.27	0.28	0.37
Primary / lower secondary education	0.30	0.44	0.34	0.00
Higher sec. educ. / interm. voc. training	0.33	0.30	0.41	0.00
Higher vocational training	0.25	0.20	0.16	0.00
University degree / university student	0.12	0.06	0.09	1.00
Female	0.45	0.53	0.56	0.46
Age 35-44 years	0.19	0.13	0.21	0.00
Age 45-54 years	0.23	0.22	0.14	0.00
Age 55-64 years	0.18	0.18	0.09	0.00
Age 65 years and older	0.16	0.33	0.10	0.00
Working	0.56	0.36	0.55	0.00
Unemployed and looking for a job	0.02	0.03	0.03	0.00
Lives in Urban Area	0.60	0.62	0.58	.
Household financial administrator	0.66	0.56	0.48	.
Employer offers savings plan	0.44	0.26	0.32	.
Has sav. plan via employer	0.36	0.17	0.26	.
Has sav. acc. or similar	0.87	0.85	0.90	.
Holds stocks, or similar	0.31	0.24	0.29	.
Net household income ∈ [22k Euros; 40k Euros)	0.49	0.51	0.49	.
Net household income at least 40k Euros	0.17	0.13	0.18	.
Maximum number of observations	1787	290	218	178

Note: The numbers shown indicate fractions in the final sample. Some households did not complete all questionnaires from the DNB Household Survey, from which some of the variables are drawn. Hence the number of observations is lower for some of the variables in question. This is particularly the case for the last two sections of the table.

in the first three columns.

2.4 Theoretical specification and empirical implementation

In order to organise the data and to interpret the magnitude of selection effects, it is important to specify a behavioural model. Here we use a simplified version of the model of von Gaudecker *et al.* (2011), assuming no preference for early or late resolution of uncertainty. Starting point

is a standard expected utility formulation with an exponential utility function:

$$(1) \qquad u(z, \gamma) \;=\; -\frac{1}{\gamma}e^{-\gamma z}$$

where $z \in \mathbb{R}$ denotes a lottery outcome and $\gamma \in \mathbb{R}$ is the Arrow-Pratt coefficient of absolute risk aversion. The choice of the exponential utility is motivated in von Gaudecker *et al.* (2011), who find that exponential utility provides a better fit of the data than than power utility.

Equation (1) is augmented with a loss aversion parameter $\lambda \in \mathbb{R}_+$, following prospect theory (Kahneman and Tversky, 1979) and in line with the widely recognised stylised fact that "losses loom larger than gains" (see, e.g., Starmer (2000) for a review):

$$(2) \qquad u(z, \gamma, \lambda) \;=\; \begin{cases} -\frac{1}{\gamma}e^{-\gamma z} & \text{for } z \geq 0 \\[2mm] \frac{\lambda-1}{\gamma} - \frac{\lambda}{\gamma}e^{-\gamma z} & \text{for } z < 0 \end{cases}$$

The degree of loss aversion is measured by the ratio of the left and the right derivatives of the utility function at zero, as suggested by Köbberling and Wakker (2005). von Gaudecker *et al.* (2011) demonstrate that this model is able to capture the main features of the data in the CentERpanel experiment and that it is superior to several alternative functional form assumptions on the grounds of model fit.[7]

Based on (2), we formulate a structural econometric model which can be estimated by maximum likelihood. The model allows for individual heterogeneity in preference parameters and in the tendency to make errors. Individual differences are captured by observed characteristics ("observed heterogeneity") or not ("unobserved heterogeneity"). Our decision problems are all choices between binary lotteries π^k, characterised by a low outcome (k^{low}), a high outcome (k^{high}) and the probability of the high outcome (p^{high}). Each pair of lotteries shares a common probability of the high outcome. Each individual $i \in \{1, \ldots, N\}$ faces $j \in \{1, \ldots, J_i\}$ dichotomous choices between two binary lotteries $\pi_j^A = \left(A_j^{low}, A_j^{high}, p_j^{high}\right)$ and $\pi_j^B = \left(B_j^{low}, B_j^{high}, p_j^{high}\right)$. Let $Y_{ij} = 1$ if the individual opts for π_j^B and $Y_{ij} = 0$ otherwise. Define the difference in certainty equivalents of the two lotteries in decision task j as:

$$\Delta \text{CE}_{ij} = \text{CE}(\pi_j^B, \gamma_i, \lambda_i) - \text{CE}(\pi_j^A, \gamma_i, \lambda_i),$$

where $\text{CE}(\pi_j^k, \gamma_i, \lambda_i)$, $k = A, B$ is the certainty equivalent of lottery π_j^k given the utility function defined by (2) with the individual-specific parameters γ_i and λ_i.

Ignoring cases of indifference (which have probability zero in our setup), a perfectly rational decision maker would choose π_j^B if and only if $\Delta \text{CE}_{ij} > 0$. As a first step to allow for stochastic decision making, we add so-called Fechner errors (see, for example, Loomes, 2005) and model the individual's choice as:

$$(3) \qquad Y_{ij} \;=\; \mathbb{I}\left\{\Delta \text{CE}_{ij} + \tau\,\varepsilon_{ij} > 0\right\},$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. We assume that the ε_{ij} are independent of each other and of the random coefficients driving the utility function, and follow a standard logistic

---

[7]In the Online Appendix, we present estimates based on several alternative models. The results show that our conclusions about selection effects are robust to the chosen functional forms and to extending model allowing Kreps and Porteus (1978) preferences towards the timing of uncertainty resolution.

distribution. Fechner errors imply that as ΔCE becomes small, the likelihood of choosing the "wrong" lottery increases. The probability of such a mistake increases with the parameter τ (> 0).

In addition to Fechner errors $\tau \varepsilon_{ij}$, we also allow for the possibility that subjects choose at random in any given task, following Harless and Camerer (1994). The propensity to do so is governed by the (individual specific) "trembling hand" parameter $\omega_i \in [0,1]$. Under these assumptions, the probability of the observed choice Y_{ij} of individual i in choice situation j, given all the individual specific parameters, is given by:

$$
\begin{aligned}
(4) \qquad l_{ij}\left(\pi_j^A, \pi_j^B, Y_{ij}, \tau, \gamma_i, \lambda_i, \omega_i\right) & = \\
& (1 - \omega_i)\,\Lambda\left((2Y_{ij} - 1)\frac{1}{\tau}\Delta\mathrm{CE}_{ij}\left(\pi_j^A, \pi_j^B, \gamma_i, \lambda_i\right)\right) + \frac{\omega_i}{2},
\end{aligned}
$$

where $\Lambda(t) = (1 + e^{-t})^{-1}$ is the logistic distribution function. For the sake of a parsimonious and easily interpretable model, we restrict τ to be the same for all individuals, but allow that subjects vary in their probability to make random choices.[8]

We use a random coefficients model for the individual-specific parameters γ_i, λ_i, and ω_i. Define

$$
(5) \qquad \eta_i = g_\eta(X_i^\eta \beta^\eta + \xi_i^\eta), \quad \eta_i \in \{\gamma_i, \lambda_i, \omega_i\}
$$

where η_i denotes one of the three individual specific parameters, X_i^η are $1 \times K^\eta$ vectors of regressors, β^η are $K^\eta \times 1$ parameter vectors, and ξ_i^η are the unobserved heterogeneity components of the parameters. The first element of each X_i^η contains 1. The functions $g_\eta(\cdot)$ are used to impose the theoretical restrictions on the individual specific parameters. This is just the identity function for γ; an exponential function ensures a positive value of λ; and the logistic distribution function guarantees that ω is always between 0 and 1. We write $g(X_i\beta + \xi_i)$ for the vector of these three functions.

We assume that the vector $\xi_i = (\xi_i^\gamma, \xi_i^\lambda, \xi_i^\omega)'$ follows a joint normal distribution independent of the regressors. von Gaudecker *et al.* (2011) find that the difference between the low and high incentive / hypothetical treatments is better captured by a multiplicative specification than by adding a low incentive dummy to X. For the low incentive treatment, we therefore multiply all slope coefficients as well as the standard deviations of the unobserved heterogeneity terms by the same parameter $\beta_{\text{low incentive}}^\eta$. Defining $\xi^* = (\Sigma')^{-1}\xi$, where $\Sigma'\Sigma$ is the covariance matrix of ξ, we can express the likelihood contribution of subject i as:

$$
(6) \qquad l_i = \int_{\mathbb{R}^3} \left[\prod_{j \in J_i} l_{ij}\left(\pi_j^A, \pi_j^B, Y_{ij}, \tau, g(X_i\beta + \xi)\right)\right] \phi(\xi^*)d\xi^*
$$

where l_{ij} is the probability given in (4) and $\phi(\cdot)$ denotes the joint standard normal probability density function with appropriate dimension (3 in this case). The log likelihood is given by the sum of the logarithms of l_i over all respondents in the sample and can be maximised by standard simulation methods.

---

[8]One might argue that τ should also be individual specific, but in practice it appears to be difficult to estimate heterogeneity in τ and ω separately (although both are identified, in theory).

3 Experimenter-induced selection

This section addresses the concern that the parameter distributions derived from standard laboratory experiments are not representative for a broader, heterogeneous, population. Our design enables analysing the effect of experimenter-induced selection while controlling for implementation modes. We first show some simple descriptive statistics highlighting the key patterns in the data before describing the results from the structural analysis.

3.1 Descriptive evidence

With regard to elicitation and modelling of preferences, the structure and frequency of errors and violations of fundamental principles of choice is of special interest. We distinguish three choice patterns that are inconsistent with almost all models of choice under uncertainty. First, a dominance violation occurs if somebody chooses option 'B' when the probability for the high outcome is zero or option 'A' when this probability is one.[9] The second type of inconsistency emerges when subjects switch back and forth from choosing 'A' and 'B' on the same screen. The third is when they switch back and forth between the initial screen and the follow-up screen (within a given payoff configuration). There was some overlap of probabilities on the two screens, so subjects could make a choice on the second that was inconsistent with their choice on the first screen.[10]

We consider the average number of violations of any of the three types as a first summary statistic for error frequencies. Only one violation per subject is counted for each payoff configuration because of the iteration scheme that depended on making an error on the first screen or not. In Figure 2, the average number of configurations with inconsistencies is presented by sample. The error frequency is twice as high in the Internet experiment (first bar) than in the laboratory experiment (second bar) and individual error propensities are significantly different using any of the standard Mann-Whitney, Kolmogorov-Smirnov, or student t-tests (two-sided p-values < 0.01). At the same time, it is evident from Figure 2 that average error frequencies are very similar across the two laboratory treatments and the Internet-Uni subsample, which consists of those respondents of the CentERpanel that are less than 35 years of age and hold a university degree or study to obtain one. The similarity between the lab and the Internet-Uni subsamples is not rejected by any of the tests. This suggests that the disparity between the lab and the Internet is not due to the different environments under which the experiments were conducted.

In Table 3, we break down the total number of inconsistencies by type of error. The pattern found in the laboratory is similar to results reported by Loomes *et al.* (2002) on a different risky choice design: Subjects make very few violations of dominance but make many more inconsistent choices when faced twice with the same decision problem. Inconsistencies between screens constitute more than 80% of all consistency violations in the Laboratory. Our results indicate that this changes dramatically when the general population is considered, where dominance violations play a much larger role and constitute 36% of all consistency violations. The numbers of within and between screens inconsistencies are also larger than

---

[9]The former can only happen on a follow-up screen if the respondent has chosen option 'B' throughout the first screen.

[10]For example, if a subject switched from 'A' to 'B' on the first screen at a 0.5 probability of the high outcome, the second screen had high outcome probabilities $0.2, 0.3, 0.4$ and 0.5. An between-screens inconsistency then arises if he or she chose 'B' at probability 0.2 or if they chose 'A' at probability 0.5.

Figure 2: Percentage of configurations with inconsistencies by sample
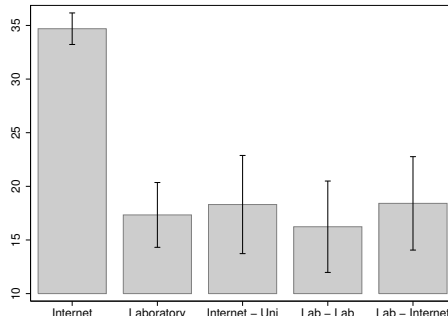
Note: The bars represent frequencies of errors as a percentage of the number of possible violations along with 95% confidence intervals. A maximum of one error was counted for each payoff configuration. "Internet" consists of unweighted numbers from CentERpanel respondents. "Laboratory" are averages for all laboratory subjects and "Internet-Uni" mean values for those respondents of the CentERpanel that are less than 35 years of age and hold a university degree or study to obtain one. "Lab-Lab" and "Lab-Internet" are averages for laboratory subjects in the "Lab-Lab" and "Lab-Internet" treatments, respectively.

in the lab. As above, the figures suggest that the difference between the laboratory and the Internet is mainly driven by differences in the sampling populations: the error frequencies of the young and well educated in the "Internet-Uni" group resemble those of the laboratory samples. The only discrepancy concerns dominance violations, which appear to be slightly more common in the "Internet-Uni" sample than in the laboratory samples.

To obtain a simple measure of risk preferences we consider at which probabilities subjects switched from (the safer) option 'A' to (the riskier) option 'B' in each payoff configuration.[11] The higher the switch point, the more averse the subject is to more risky choices. Similar measures have been used in earlier studies, cf., e.g., Holt and Laury (2002). The average switch point of 70% in the Internet experiment is considerably higher than the corresponding figure of 61% for the laboratory experiment. The findings in the laboratory are close to those in Holt and Laury (2002) for the payoffs most comparable to ours. Figure 3 confirms that similar difference between the two samples are found for all seven decision problems. Using the MW and KS tests we find that the differences between the laboratory and Internet samples are highly significant, both comparing averages across all questions and looking at each question separately (two-sided p-values < 0.01). For the sake of brevity, we do not report separate results on the Internet-Uni subsample, or on differences between the two mode treatments in the Laboratory. They show the same pattern we found for the inconsistencies: Controlling for

---

[11]We can only compute bounds that will at best be a 5%-interval (e.g. between 75% as the highest 'A'-choice on the first screen and 80% as the lowest choice of 'B' on the second screen). In many cases, inconsistencies make these bounds grow substantially wider. We computed them as follows: the lowest possible switch point is defined as the highest probability corresponding to an 'A' choice that is still lower than the minimum probability with a 'B' choice; the upper bound is the minimum probability with a 'B' choice that is still higher than the maximum probability where option 'A' was chosen. If only choice 'A' ('B') was observed, both upper and lower bound were set to 100% (0%). We then averaged the upper bounds and the lower bounds across the seven payoff configurations. This leaves us with two preference measures per individual, To save space, we just report results using the midpoint of the two bounds. All results remain qualitatively the same if we use the upper or lower bounds, figures corresponding to Figure 3 can be found in the Online Appendix.

Table 3: Relative frequency of inconsistencies by type of error and subsample

Sample	# Obs.	Dominance	Within	Between	Total
Internet	1787	12.6%	4.9%	24.5%	34.7%
Laboratory	178	1.5%	2.0%	14.6%	17.3%
Internet - Uni	96	4.2%	1.9%	14.4%	18.3%
Lab - Lab	88	1.6%	1.9%	13.9%	16.2%
Lab - Internet	90	1.5%	2.0%	15.2%	18.4%

Note: The figures represent frequencies of the different types of errors as a percentage of the number of possible violations. The fractions of violations for the dominance category were obtained by dividing the total number of dominance violations in each category by the total number of screens shown to each subject on which dominance violations could be made. The numbers for the within category are calculated as the number of within violations, divided by the total number of screens shown to each subject. The figures of the last column were obtained by dividing the number of between errors by the number of times the second screen was displayed to subjects. A maximum of one error was counted for each payoff configuration.

the characteristics of the sampling population eliminates the differences between the Internet and the Laboratory, mode effects appear to play no role.

Figure 3: Mean switch points in the CentERpanel and laboratory experiments
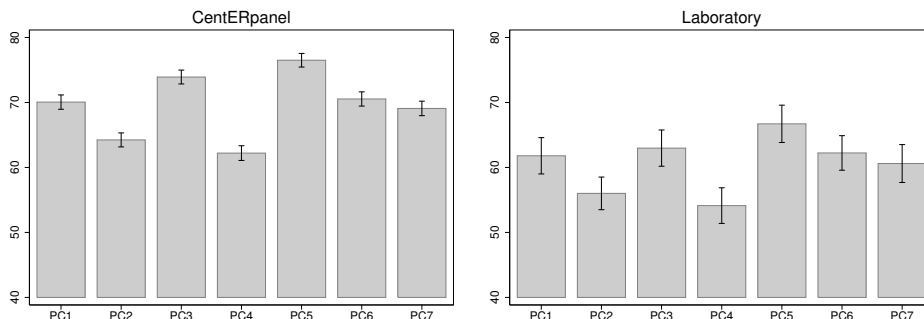
Note: The numbering of the payoff configurations (PC) conforms to those in Table 1. "Switch points" are defined as the mean of the 'lowest' and 'highest' switch points discussed in the text and depicted in the Online Appendix, Figures 7 and 8, respectively. Error bars depict 95% confidence intervals.

3.2 Results from the structural model

The descriptive results are suggestive for differences in mean parameters, but difficult to interpret in magnitude. This applies even more to higher moments like the variance, which is why we have not discussed them so far. Error propensities are underestimated because violations of revealed preference conditions across configurations cannot be incorporated. Estimates of the structural model described in Section 2.4 overcome these limitations. Table 4 displays the estimation results using the combined sample including subjects from both the

14

laboratory and Internet experiments. In the laboratory, subjects received a participation fee irrespective of treatment, whereas only subjects in the incentive treatments received a participation fee in the Internet experiment. This creates a small discrepancy between the laboratory and the Internet experiments with respect to the hypothetical treatment. As a precautionary measure, the results presented in Table 4 are therefore based only on the observations from the two incentive treatments. This restriction on the sample does not matter for our conclusions, which can be verified by consulting the table based on the full sample available in the Online Appendix.[12]

In order to facilitate the interpretation of results, the values in Table 4 have been transformed to the original parameter scale – the constant terms are simply $g_\eta(\beta_1^\eta)$. The first element in each column therefore represents the median (preference or error) parameter in the high incentive treatment. For the Internet-Uni and Laboratory coefficients, the table shows the partial effect on the median parameter value of setting the variable to one given the reference value defined by the left-out category: $g_\eta(\beta_{\text{constant}}^\eta - \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$. Since the low incentive treatment enters multiplicatively, we report the non-transformed parameter, $\beta_{\text{low incentive}}^\eta$; values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. We only report the results imposing a diagonal structure on Σ; substantive results for a general variance-covariance matrix are very similar and reported in the Online appendix.

The estimation results in Table 4 reveal significantly lower median parameter values for risk aversion, loss aversion, and error propensities among the laboratory participants. The findings for γ and λ mean that risk-taking is considerably lower among CentERpanel respondents than in the laboratory. This is in line with Figure 3, but contrasts with the finding by Andersen *et al.* (2010) of similar average parameters among students and the general Danish population. The most likely explanation lies in greater power of our estimates with 1790 (178) individuals in the CentERpanel (lab) experiments compared to 253 (90) subjects in Andersen *et al.*'s field (lab) experiments. Median random choice propensities are spectacularly lower in the laboratory than in the CentERpanel experiment. This is in line with the low error propensities found by Hey and Orme (1994) or Loomes *et al.* (2002) in laboratory experiments; and the high rates estimated by de Roos and Sarafidis (2010) or Huck and Müller (2007) for non-student samples.

The similar coefficients in the second and third row of Table 4 indicate once more that subjects in the Internet-Uni subsample display behaviour similar to the laboratory subjects. This is confirmed by a likelihood ratio test between the specification underlying Table 4 and an alternative specification with a single dummy variable for the laboratory and Internet-Uni samples combined (p-value= 0.24). In line with the descriptive evidence of the previous section, we do not find any significant differences between the Lab-Lab and the Lab-Internet treatments using the structural model. Consequently, in the presentation of the results we have chosen to merge the two lab treatments (see the Online Appendix for estimation results distinguishing between these two treatment groups). We conclude that implementation mode effects do not matter. By concentrating on student samples, typical experiments will end up with a selection of experts who are less risk averse and loss averse than the general population.

Average parameters are only part of the story. In order to study how the distributions

---

[12]We do observe that the hypothetical treatment affects loss aversion estimates differently in the laboratory compared to the Internet, which is likely to be caused by the presence of participation fees in the laboratory. The results concerning selection effects are robust to the inclusion of the hypothetical treatment.

Table 4: Parameter estimates from the preferred specification based on utility function (2), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0360***	2.25***	0.0788***
	(0.0012)	(0.106)	(0.0084)
Internet university subsample	-0.0174***	-0.305	-0.0557***
	(0.0036)	(0.356)	(0.0094)
Participant in Lab experiment	-0.0186***	-0.765***	-0.0687***
	(0.0043)	(0.270)	(0.0074)
Low incentive treatment[†]	2.57***	1.08***	1.24
	(0.0747)	(0.0324)	(0.158)
σ	0.0376***	1.19***	2.07***
	(0.0008)	(0.0404)	(0.113)

Note The value of the log-likelihood function is 28171.4, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 4.16 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

differ, we estimate separate models for the CentERpanel and laboratory experiments. In Figure 4, we plot the parameters' cumulative distribution functions. All parameter distributions exhibit much more dispersion when estimated for the Internet experiment than for the laboratory experiment. For the risk aversion parameter, this mirrors the finding by Andersen *et al.* (2010). The visually apparent difference is corroborated by likelihood ratio tests comparing the separate specifications for the two experiments and the joint specification underlying Table 4 (p-values < 0.01).

Considering the quantitative importance of the differences in heterogeneity, the 5% and 95% quantiles of the distribution of γ for the laboratory subjects are $[-0.02, 0.05]$ compared to $[-0.036, 0.1]$ for the Internet subjects. Similarly, the same interval for the λ parameter in the laboratory, $[0.5, 5.5]$, is considerably smaller than the corresponding interval for the Internet experiment $[0.28, 16]$, where the lower bounds are artefacts of the functional form of the distribution (see von Gaudecker *et al.*, 2011). Moreover, only 8% of subjects in the student sample have $\lambda > 5$, compared to 25% in the Internet sample. The heterogeneity in error propensities is also much smaller in the laboratory experiment, where 90% of subjects are estimated to have a random choice probability in the interval $[< 0.001, 0.38]$ compared to $[0.02, 0.7]$ for the Internet sample. Using laboratory samples alone will considerably underestimate heterogeneity in all dimensions.

Figure 4: Distributions of preference and error parameters in the CentERpanel and Laboratory experiments
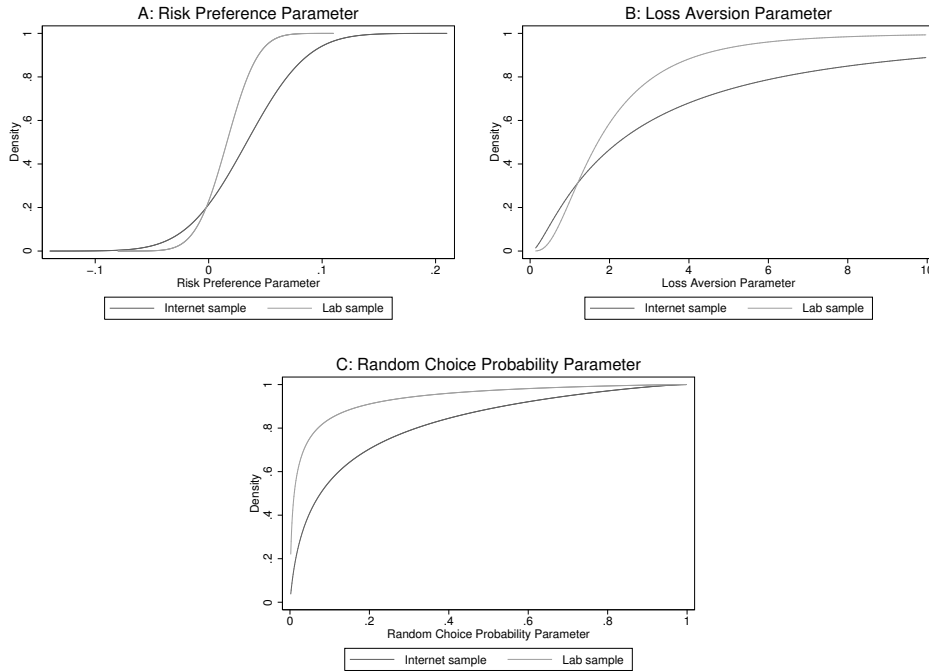
Note: Graphs are based on separate estimations for the CentERpanel and Laboratory samples, using the real incentive treatments only, a diagonal covariance matrix, and a minimal set of covariates (coefficients can be found in the Online Appendix, Tables 15 and 16)).

4 Self-Selection Bias in the CentERpanel Experiment

The results of the previous section highlight that the composition of the sample can have a significant impact on the estimated distributions of relevant parameters. Moving from the laboratory to the Internet survey clearly enhances the heterogeneity and representativeness of the sample. Yet, in order to get reliable population-wide estimates, sampling from a representative sub-population suffices only if non-response is perfectly random. Since people self-select into the experiment this condition may not hold. Indeed, the descriptive statistics in Table 2 suggest that there are some important differences between those who completed the experiment without rushing through it, those who chose not to participate, and those who started but rushed through or did not complete in terms of observed background characteristics. The question is whether this also leads to selection effects in the estimates of risk preferences or the tendency to make inconsistent decisions.

A few existing studies address similar topics. Bellemare and Kröger (2007) use the same sampling frame as we do and find no evidence of selective non-participation for measuring trust. On the other hand, Harrison *et al.* (2009) conclude that self-selection effects are important for estimated risk preferences; their sampling frame and recruitment procedures substantially differ from ours, however.

We first analyse the determinants of self-selection and then investigate their impact on observed choices. In order to structure the analysis, it is useful to divide the sampling process in the Internet experiment into three stages:

1. Dutch individuals are contacted at random and participate in the CentERpanel or not.

2. A random subsample of CentERpanel respondents is asked to take part in our experiment. After learning about the nature of the experiment, they decide to participate or decline participation.

3. Some of the subjects drop out during the experiment or click through it extremely rapidly.

To see how step 2 relates to typical recruitment procedures, note that some information on payoffs and the type of experiment is usually conveyed in recruitment emails, or on posters announcing the experiment, before subjects come to the lab. Such information is provided on our welcome screen (Figure 6 in the Online Appendix). Subjects learn about the nature of the experiment and the possible payoffs, and then choose to participate or not.

Dropping out or not taking the task seriously usually does not play a role in the laboratory. Hence, step 3 seems typical for the Internet environment. However, we think that participation decisions for laboratory experiments combine features of steps 2 and 3 because of the negligible fixed costs of participation in the CentERpanel experiment. Showing up at the laboratory at a specific time and date entails a significant cost – and subjects can be expected to have made the trade-off between the costs and benefits of participation beforehand. This is probably not the case in the Internet setting, where the experiment can be accessed within seconds of notification. Hence the cost-benefit analysis may well be postponed and carried out during the experiment. This may explain why subjects hardly ever leave the economics laboratory prematurely (and nobody left our laboratory sessions), whereas 4% of subjects did not finish the CentERpanel experiment. Similarly, rushing through the experiment can be considered as a form of non-participation, since there is a lower bound on the time needed

to digest the instructions and to give serious answers. This minimum time certainly seems higher than the 1:43 minutes which is the minimum time observed in the Internet experiment. We checked several cut-off points between 3 and 7 minutes and finally chose the minimum duration observed in the laboratory (5:20 minutes). Results were robust to the precise value chosen. With this threshold, about seven percent of the Internet subjects fall into the category of "speeders."[13]

An alternative explanation why step 3 features prominently in the Internet experiment and not in laboratory experiments is the interaction with the experimenter and the typical rules in the laboratory. One difference is the possibility to ask questions. Internet participants who do not understand a task and cannot ask questions may more easily opt for randomly ticking options or drop out entirely. Another difference is that in typical laboratory experiments everybody is expected to stay until the last subject has finished, so that there is no point in rapid completion. We can analyse the consequences of these differences by comparing the "Lab-Lab" and "Lab-Internet" treatments (see Section 2.3). The distributions of the completion times are similar, with mean durations of about 12.5 minutes in both cases. Surprisingly, in the traditional "Lab-Lab" treatment the dispersion is higher and the left tail of the distribution has more mass. If rapid completion were due to the two factors mentioned above, we would expect the opposite. Completion times in the "Internet-Uni"- subgroup of the Internet subjects are lower than those of the laboratory subjects. This is consistent with our preferred interpretation of step 3 of the selection process, since the "Internet-Uni" group will contain more respondents who rush through the experiment.

4.1 The determinants of self-selection

To analyse the factors that drive participation in the Internet experiment, we estimated a multinomial logit model with three possible outcomes: non-participation, rushing through or dropping out, and regular (full) participation. Results are presented in Table 5, with full participation as the baseline category. The regressors are taken from the variables that were also used in Table 2. Columns 1 and 2 contain the coefficients and standard errors of our basic specification, for non-participation and dropouts/speeders, respectively. Only basic covariates that are available for almost everybody are included – dummies for the incentive treatments, gender, education, age and occupational status. In the extended specification (columns 3 and 4) the number of observations is lower since we use covariates which are not available for all subjects. This specification adds household income (in three categories), household wealth (four categories) and variables measuring financial expertise and preferences.

For the variables included in both specifications, results for the two sets of estimates are very similar. Non-participation is significantly less likely in the incentive treatments than in the hypothetical treatment (the benchmark). Translated into marginal effects, the coefficients indicate response rates that are almost ten percentage points higher than in the hypothetical treatment (all marginal effects are evaluated at a baseline with all dummy variables set to zero). The point estimates on not finishing the experiment are much smaller in magnitude and not significant. While incentives increase participation, they do not seem to attract

---

Table 5: Self-selection into the CentERpanel experiment

Specification	NP (1)	DO/SP (2)	NP (3)	DO/SP (4)
Low incentive treatment	-1.043*** (.163)	-.219 (.174)	-1.103*** (.193)	-.310 (.215)
High incentive treatment	-.699*** (.156)	-.277 (.183)	-.847*** (.195)	-.318 (.231)
Female	.388*** (.137)	.273* (.152)	.393** (.165)	.222 (.190)
Primary / lower secondary education	.853*** (.280)	.463* (.277)	.741** (.362)	.599* (.357)
Higher sec. educ. / interm. voc. training	.580** (.285)	.337 (.267)	.571 (.360)	.255 (.346)
Higher vocational training	.458 (.294)	-.222 (.299)	.442 (.368)	-.174 (.372)
Age 35-44 years	.335 (.242)	-.449** (.196)	.373 (.311)	-.502* (.264)
Age 45-54 years	.578*** (.221)	-1.075*** (.222)	.706** (.283)	-1.077*** (.287)
Age 55-64 years	.486** (.231)	-1.321*** (.262)	.737** (.289)	-1.283*** (.319)
Age 65 years and older	1.137*** (.232)	-1.179*** (.278)	1.353*** (.303)	-1.425*** (.359)
Working	-.390** (.174)	-.111 (.179)	-.046 (.224)	.006 (.240)
Unemployed and looking for a job	.068 (.416)	.188 (.438)	.125 (.524)	.109 (.573)
Net household income $\in$ [22k Euros; 40k Euros)			.163 (.188)	.132 (.217)
Net household income at least 40k Euros			-.022 (.276)	.481 (.293)
Assets worth $\in$ [10k Euros, 50k Euros)			-.168 (.280)	.409 (.275)
Assets worth $\in$ [50k Euros, 200k Euros)			.216 (.214)	.070 (.252)
Assets worth more than 200k Euros			.207 (.241)	.347 (.293)
Household financial administrator			-.353** (.172)	-.293 (.202)
Employer offers savings plan			.107 (.296)	-.590 (.401)
Has sav. plan via employer			-.830*** (.321)	.114 (.413)
Has sav. acc. or similar			.163 (.346)	.177 (.413)
Holds stocks, or similar			-.410** (.184)	.007 (.210)
Constant	-2.467*** (.340)	-1.634*** (.321)	-2.551*** (.570)	-1.864*** (.597)
Number of observations	2295	2295	1690	1690

Note: Coefficient estimates and corresponding standard errors of multinomial logit regression. Columns indicate categories of the dependent variable by regression type. Columns (1) and (3) list estimates for opting for non-participation on the first screen (NP); columns (2) and (4) those for dropping out before completion (DO) or finishing the experiment in less than 5:20 minutes (SP). Asterisks indicate significance at the 10%, 5%, and 1%-level.

a systematically different group of participants: we estimated models that included a wide variety of interaction effects and none of these was significant. The coefficients on the high and low incentive treatments are not significantly different from each other.

Persons in the top two education categories are both significantly more likely to participate in the experiment and to finish it. Women's non-participation rates are four to five percentage points higher than those of men. Women also are slightly more prone to quit during the experiment or to finish it very rapidly. Age effects start to matter at age 45, beyond which participation rates decline. Those beyond 65 years of age are only half as likely to start the experiment as those younger than 35. At the same time, however, non-completion rates decrease significantly with age, mainly due to the fact that older participants are less likely to rush through the experiment.[14] The combined effects of age on full participation are small and insignificant in almost all cases. Working respondents have higher participation rates than non-workers according to the parsimonious specification, but the effect becomes insignificant in the richer specification. Labour market status does not affect quitting or speeding.

Point estimates of the effects of income and wealth on participation are generally small and insignificant, and a joint test does not reject the null hypothesis that they play no role. The other financial variables are proxies for preferences and financial knowledge. Being the financial administrator of the household, which may reflect a preference for spending one's time with problems involving financial choices and taking financial risks, significantly increases the propensities to participate. A dummy for whether the employer offers a savings plan is a control variable necessary to avoid confounding the effects of holding such a plan with employment type. The variable of interest, taking part in a save-as-you-earn savings arrangement, is associated with a (significant) six percentage points higher propensity to begin the experiment, supporting the interpretation that participation in the experiment is associated with financial expertise and interest. This interpretation is strengthened by the other two portfolio variables: On the one hand, having an ordinary savings account does not have any predictive power for taking part in the experiment, since saving accounts are commonly known and do not require much expertise or effort; on the other hand, the ownership of mutual funds, stocks, etc., is significantly associated with higher participation. These are much more sophisticated products and investing in them requires some financial knowledge. The results thus suggest that interested and knowledgeable individuals have a higher probability of participating.

4.2 The impact of self-selection on outcomes

To test whether self-selection effects matter for preferences and error propensities, we compare the observed (unweighted) sample distribution of full participants with a weighted distribution that corrects for the various steps of the selection process. For step 1, CentERdata provides standard survey weights based upon comparing with a much larger household survey drawn by Statistics Netherlands. We will assume that selection into the CentERpanel is independent of the variables of interest, conditional on the basic background variables used to construct these weights (age, sex, education, home ownership, region). This is a missing at random (MAR) assumption, see e.g. Little and Rubin (2002)). It implies that the standard weights

---

[14]We estimated models that treated the two components of step 3 (speeding through and dropping out) as separate outcomes. Age is the only variable for which the distinction mattered, which is why we only report the results from the more parsimonious specification in the table.

can be used to correct for the selection in step 1. We make similar MAR assumptions for the other two steps, but then conditioning on the much larger set of background variables used in the previous subsection. We construct weights from a probit model that jointly explains the selection in steps 2 and 3; each weight is the inverse of the predicted probability of being in the final sample. We multiply these weights with the weights for step 1 to get weights that correct for all three steps of the selection process. Individual estimates of the preference and error parameters are obtained from the structural model using an approach similar to that of Revelt and Train (2000). The ("posterior") distributions of the parameters of each respondent are derived, conditional on observed choices of that respondent and given the estimated ("prior") distributions of the parameters. We then use the median of each individual's parameter distribution as the unit of observation and test whether the estimated parameters of the weighted sample are significantly different from unweighted estimates. This can be seen as a test whether the selection process is selective, under the maintained assumption that selection in each step is MAR given all the covariates used to construct the weights. For this purpose we include a minimal set of covariates in the structural estimation, consisting only of the treatment dummies. We opt for the full specification in the probit regression, similar findings emerge for the parsimonious set of covariates, see Section 18 in the Online Appendix.

In Table 6 the average preference and error parameters for the weighted and unweighted data are presented together with p-values of t-tests comparing mean values.[15] Taking the full selection process (steps 1, 2 and 3) into account, we observe no effect on risk aversion and a significant, albeit small, effect on loss aversion, with self-selection leading to a slightly less loss averse sample. The effects are very small compared to experimenter-induced selection and plotting the weighted distribution against the unweighted one hardly reveals visual differences (Figure 5). Table 6 reveals that most of this effect originates from steps 2 and 3, while step 1 has less of an impact.

The largest effect of self-selection relates to the tendency to make random errors captured by ω. The unweighted mean parameter is 0.214 compared to 0.236 when taking the full selection process into account. The differences between the two weighted estimates and the unweighted are highly significant (p-values:< 0.001), and, in contrast to the cases of risk and loss attitudes, also the difference between the two weighted averages is borderline significant (p-values:$= 0.098$). Still, the selection effect is mostly driven by steps 2 and 3. Figure 5 demonstrates the larger impact of self-selection on the distribution of error propensities than for the preference parameters. However, it is clearly of secondary importance compared to the experimenter-induced selection effects depicted in Figure 4. Thus, although the participation decision is correlated with demographics, selection on observables hardly alters the estimates of risk preferences and leads to a moderate underestimation of the error propensity. The sampling procedure still leads to expert subjects, but the self-selection effect seems much less pronounced than the effect caused by a narrow sampling population.

Using the approach developed by Heckman (1976, 1979), Harrison *et al.* (2009) find that not controlling for selection effects leads to an overestimation of average risk aversion in the population. The structural model, combined with the MAR assumptions, does not allow for

---

[15]The test works as follows: Let y denote our parameter of interest and $w(x)$ the weight. Our null hypothesis of no difference between the weighted and unweighted observations can then be stated as $E[w(x)y] = E[y]$ or $E[z] = 0$, with $z = (w(x) - 1)y$. Since we have a large sample size and few explanatory variables, we neglect the estimation error in w. The null hypothesis can then be tested with a standard t-test on whether the mean of z is zero or not.

Table 6: Significance of selection weights, loss aversion utility specification (2), diagonal covariance matrix,

Parameter / Weight	Average	P-value, steps 2, 3	P-value, no weight
γ – Weights for steps 1 to 3	0.0520 (0.0686)	0.570	0.248
γ – Weights for steps 2 and 3	0.0515 (0.0696)	.	0.036
γ – Unweighted	0.0539 (0.0743)	.	.
λ – Weights for steps 1 to 3	11.52 (30.59)	0.744	0.022
λ – Weights for steps 2 and 3	11.42 (31.36)	.	0.011
λ – Unweighted	10.53 (28.63)	.	.
ω – Weights for steps 1 to 3	0.236 (0.249)	0.098	0.001
ω – Weights for steps 2 and 3	0.230 (0.245)	.	0.000
ω – Unweighted	0.214 (0.237)	.	.

Note: Treatment effects are netted out. Standard errors are given in parenthesis. P-values are based on t-tests, described in Footnote 15 in Section 4.2, with the null hypotheses of equal means.

a direct comparison with their findings. We therefore also estimated a model very similar to the one of Harrison *et al.* (using the same exclusion restriction regional variables affect selection but not risk preferences). We take the individual parameter estimates previously used in Table 6 as dependent variables. The results are displayed in Tables 20-25 of the Online appendix and they corroborate our earlier findings: Regardless of whether we condition on the broad or narrow set of covariates in Table 5 the selection effect is insignificant across all parameters. However, for the error parameter, the effect of selection is borderline significant (p-values .109 and 0.100). This reinforces the earlier conclusion that in our experimental set up, self-selection had the strongest impact on the tendency to behave in a consistent manner and had little effect on estimated risk preferences.

Figure 5: Distributions of preferences and errors in the CentERpanel experiments by selection adjustment
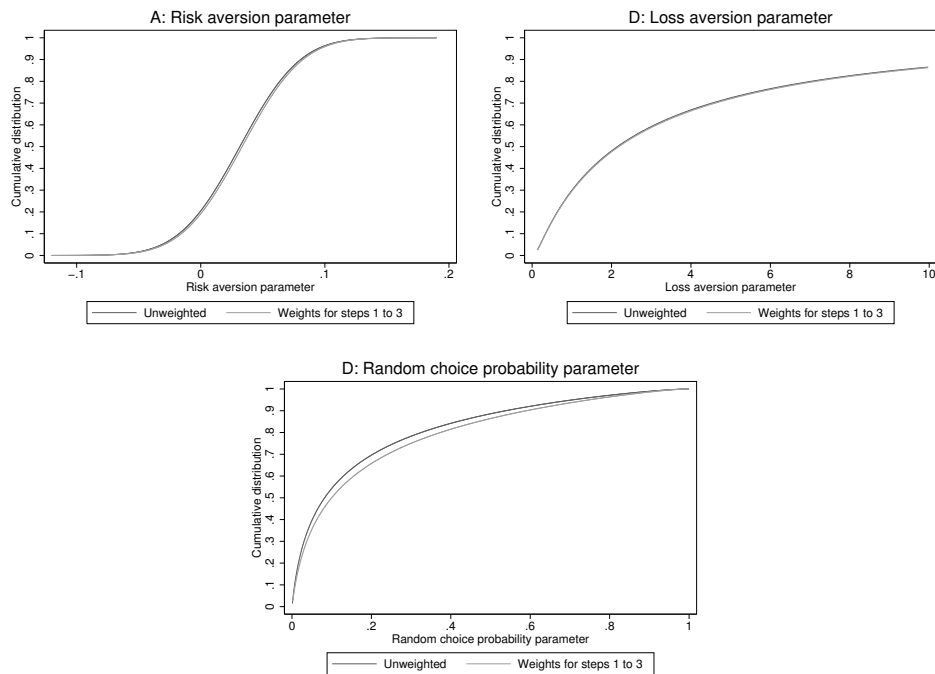
Note: Graphs are based estimations CentERpanel sample, using the full set of covariates and a diagonal covariance matrix. Coefficients can be found in the Online Appendix, Table 19. The weights are based on all three selection steps, i.e. the sampling weights provided by CentERdata and the results from a probit equation predicting drop-out in steps 2 or 3, using the same covariates as in Table 19.

5 Conclusions

We have analysed two potential sources of selection bias in economic experiments. First, we have addressed sample selection effects generated by employing a demographically homogenous subject group of university students. Second, we have investigated the importance of self-selection into experiments with subjects drawn randomly from the broad population, using detailed background information on both participants and non-participants. These aims were achieved by utilising data from both a typical laboratory experiment and an Internet experiment with subjects drawn from the complete Dutch population of 16 years and older.

Our results show that subjects in the laboratory are substantially less risk averse and loss averse than the average adult individual in the Dutch population. The laboratory subjects have strikingly lower error rates. Moreover, in line with the study of Andersen *et al.* (2010), we find that the behaviour of the student sample in the laboratory is relatively uniform, failing to reproduce the wide-ranging preference heterogeneity of the general population.

Laboratory experiments serve as useful testbeds for evaluating different methods of preference elicitation, and estimating causal effects of treatment variables (Falk and Heckman, 2009). Our results indicate, however, that laboratory experiments are less helpful for calibrating risk preference parameters for population groups not represented in the laboratory. For example, the loss aversion parameter is often taken to be $\lambda \in [2, 2.5]$ (e.g. Benartzi and Thaler, 1995) based on medians in laboratory experiments. Our analysis shows that this is likely to be a strong underestimation for the average parameter in the complete adult population. Considering estimates of preference heterogeneity, the laboratory experiment did an even poorer job: heterogeneity among laboratory subjects is a lot smaller than heterogeneity of risk preferences in the general population. Taken together, this also suggests that it is problematic to extrapolate findings about risk preferences in the lab to more general populations.

To solve the selection problem with students in the lab, subjects in an experiment can be selected on the basis of a probability sample drawn from the population of interest. One way in which this can be done is to embed the experiment into an existing socio-economic survey. The second goal of our study is to investigate the feasibility of this approach for measuring risk preferences, implementing the same experiment in an existing representative Internet survey. In principle, the several steps in selecting subjects with this approach can still induce selection bias, since participation in the ongoing survey as well as the specific experiment is voluntary. Our results suggest, however, that these selection effects are negligible where it comes to measuring the distribution of risk aversion and loss aversion. Average measures of risk aversion and loss aversion were nearly unaffected by self-selection. We did find a selection effect on error rates – error rates were reduced by around 20% due to self-selection of the participants in the Internet survey. These conclusions are obtained under the assumption that selection effects can only enter through observables, but remain unchanged when allowing for selection through unobservables using a Heckman selection model.

Moving from the traditional experimental laboratory inevitably changes the environment under which the experiment is conducted and one might fear that the loss of control on the Internet would affect behaviour. We addressed this issue by running laboratory treatments mimicking the Internet experiments as well as analysing a subsample of young and highly educated subjects in the Internet experiment. Both approaches lead to the conclusion that running the experiments over the Internet did not lead to significantly different results com-

pared to running the experiments in a more standard laboratory setting. In other words, we did not find effects of implementation mode on estimated risk preferences.

All in all, these results support the conclusion that embedding the experiment in a representative Internet survey has the potential to overcome the selection problems with student samples in the lab. Self-selection due to voluntary participation in the Internet experiment appears to be much less of a problem than the experimenter-induced restriction to convenience (student) samples. However, the high error rates observed in the subjects from the general population in Internet experiment suggest a need to take special care when designing experiments aimed at a broad population; implementing graphical aids as in **?** or our application, help screens and other instructional devices not commonplace in laboratory experiments may provide invaluable when moving outside of the laboratory. For the relatively simple decision problems presented to the subjects in our experiment, this seems to be feasible.

References

ALESSIE, R., HOCHGÜRTEL, S. and VAN SOEST, A. (2006). Non-take-up of tax-favored savings plans: Evidence from dutch employees. *Journal of Economic Psychology*, **27** (4), 483–501.

ANDERSEN, S., HARRISON, G. W., LAU, M. I. and RUTSTRÖM, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, **9** (4), 383–405.

—, —, — and — (2008). Eliciting risk and time preferences. *Econometrica*, **76** (3), 583–618.

—, —, — and — (2010). Preference heterogeneity in experiments: Comparing the field and laboratory. *Journal of Economic Behavior & Organization*, **73** (2), 209–224.

BELLEMARE, C. and KRÖGER, S. (2007). On representative social capital. *European Economic Review*, **51** (1), 183–202.

—, — and VAN SOEST, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, **76** (4), 815–839.

BENARTZI, S. and THALER, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, **110** (1), 73–92.

BINSWANGER, H. P. (1980). Attitudes towards risk: An experimental measurement in rural india. *American Journal of Agricultural Economics*, **62**, 395–407.

BLEICHRODT, H., PINTO, J. L. and WAKKER, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, **47** (11), 1498–1514.

BLUNDELL, R. W. and STOKER, T. M. (2007). Models of aggregate economic relationships that account for heterogeneity. In J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics*, Handbook of Econometrics, vol. 6, Part 1, *68*, Elsevier, pp. 4609 – 4666.

BROWNING, M., HANSEN, L. P. and HECKMAN, J. J. (1999). Micro data and general equilibrium models. In J. B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, vol. 1, *8*, Elsevier, pp. 543–633.

BWGRID (2007-2010). `http://www.bw-grid.de/`, Member of the German D-Grid initiative, funded by the Ministry of Education and Research (Bundesministerium für Bildung und Forschung) and the Ministry for Science, Research and Arts Baden-Württemberg (Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg).

CHOI, S., FISMAN, R., GALE, D. and KARIV, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, **97** (5), 1921–1938.

DE ROOS, N. and SARAFIDIS, Y. (2010). Decision making under risk in deal or no deal. *Journal of Applied Econometrics*, **25** (6), 987–1027.

DOHMEN, T., FALK, A., HUFFMAN, D., SUNDE, U., SCHUPP, J. and WAGNER, G. G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey, iZA Discussion Paper No. 1730.

—, —, —, —, — and — (forthcoming). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*.

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.

FALK, A. and HECKMAN, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, **326** (5952), 535–538.

FEHR, E., FISCHBACHER, U., VON ROSENBLADT, B., SCHUPP, J. and WAGNER, G. G. (2003). A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative surveys, iZA Discussion Paper No. 715.

GÜTH, W., SCHMIDT, C. and SUTTER, M. (2007). Bargaining outside the lab – a newspaper experiment of a three person-ultimatum game. *Economic Journal*, **117** (518), 449–469.

HARLESS, D. W. and CAMERER, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, **62** (6), 1251–1289.

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

—, — and — (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization*, **70** (3), 498–507.

—, — and WILLIAMS, M. B. (2002). Estimating discount rates in denmark: A field experiment. *American Economic Review*, **92** (5), 1606–1617.

— and LIST, J. A. (2004). Field experiments. *Journal of Economic Literature*, **42** (4), 1009–1055.

—, — and TOWE, C. (2007b). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, **75** (2), 433–458.

HECKMAN, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, **42** (4), 679–94.

— (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, **5** (4), 475–492.

— (1979). Sample selection bias as a specification error. *Econometrica*, **47** (1), 153–61.

HEY, J. D. and ORME, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, **62** (6), 1291–1326.

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

HUCK, S. and MÜLLER, W. (2007). Allais for all: Revisiting the paradox, mimeo, University of Tilburg. Available at: http://center.uvt.nl/staff/muller/Allais071009.pdf.

KAHNEMAN, D. and TVERSKY, A. V. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.

KÖBBERLING, V. and WAKKER, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, **122**, 119–131.

KREPS, D. M. and PORTEUS, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, **46**, 185–200.

LAZEAR, E., MALMENDIER, U. and WEBER, R. (2011). Sorting, prices, and social preferences, available at `http://www.econ.berkeley.edu/~ulrike/research.html`.

LEVITT, S. D. and LIST, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, **21** (2), 153–174.

LITTLE, R. J. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons Inc., 2nd edn.

LOOMES, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, **8** (4), 301–323.

—, MOFFATT, P. G. and SUGDEN, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, **24** (2), 103–130.

LUCKING-REILEY, D. (1999). Using field experiments to test equivalence between auction formats: Magic on the internet. *American Economic Review*, **89** (5), 1063–1080.

MADRIAN, B. C. and SHEA, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, **116** (4), 1149–1187.

REVELT, D. and TRAIN, K. E. (2000). Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier, university of California at Berkeley, Economics Working Paper E00-274.

STARMER, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, **38** (2), 332–382.

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), 557–71.

Tversky, A. V. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5** (4), 297–323.

von Gaudecker, H.-M., van Soest, A. and Wengström, E. (2008). Selection and mode effects in risk preference elicitation experiments, iZA Discussion Paper No. 3321.

—, — and Wengström, E. (2011). Heterogeneity in risky choice behaviour in a broad population. *American Economic Review*, **101** (2).

Additional Material to Accompany:

Experts in experiments:
How selection matters for estimated distributions of preferences

Hans-Martin von Gaudecker
Arthur van Soest
Erik Wengström

March 12, 2011

6 Alternative utility specifications: Theoretical framework

In this section, we provide some evidence on the importance of the features of the utility function in the paper. First, we simply use the constant absolute risk aversion utility specification (1), i.e. we restrict λ to be one for everybody. The second function moves us closer to prospect theory (Kahneman and Tversky, 1979). The valuation part of cumulative prospect theory (Tversky and Kahneman, 1992) has two main ingredients: A utility function with a reference point and probability weighting. We do not model probability weighting because of the reasons given in von Gaudecker *et al.* (2011). The utility function has three main characteristics: It is convex for losses (i.e. outcomes smaller than the reference point), concave for gains, and steeper for losses than for gains (this notion has been made precise by Köbberling and Wakker (2005)). This formulation has been shown to yield a useful description of mean or median behaviour in a variety of studies (Tversky and Kahneman, 1992).

We noted in Section 2.4 that our data do not permit us to estimate separate utility curvature parameters for gains and losses. Invoking an assumption of switching risk preferences is straightforward for individuals with concave utility curvature on the positive domain. However, risk loving for purely positive gambles is important for up to 20% of the population and prospect theory is silent on such behaviour. Restricting γ to be larger than zero on the positive domain led to numerical difficulties because the optimisation routine put a lot of mass as close to zero as possible in trying to move below zero. The most natural assumption for us was to take these persons to have the same value of γ on the entire real line, as opposed to assuming switching risk preferences. Equation (1) is then modified to be:

$$
(7) \qquad u(z,\gamma,\lambda) \;=\; \begin{cases} \frac{1}{\gamma} - \frac{1}{\gamma}e^{-\gamma z} & \text{for } z \geq 0 \\[2mm] \frac{\lambda}{\gamma} - \frac{\lambda}{\gamma}e^{-\gamma z} & \text{for } z < 0 \wedge \gamma < 0 \\[2mm] -\frac{\lambda}{\gamma} + \frac{\lambda}{\gamma}e^{\gamma z} & \text{for } z < 0 \wedge \gamma > 0 \end{cases}
$$

The third specification takes into account attitudes towards the resolution of uncertainty. We follow von Gaudecker *et al.* (2011) and adopt the general framework of Kreps and Porteus (1978). In line with our experimental setup we consider a two-period setting. All decisions are made in the first period and payments are made in the second period. The outcome of a gamble is either revealed in period 1, directly after all choices have been made (early resolution), or at the time of the payments in period 2 (late resolution). Assume that agents first calculate period 2 utility for all outcomes based on a function $v(z,\cdot)$, where z is the payoff and the dot replaces the preference parameters. Thereafter, agents are assumed to use a continuous and strictly increasing weighting function $h(\cdot)$ to calculate their first period utility, with period 2 utility v as its argument. The period 1 utility of a degenerate lottery that gives a certain outcome in period 2 is then given by $h(v(z,\cdot))$. The evaluation of nondegenerate lotteries hinges on the timing of uncertainty resolution: let $V(\pi)$ denote the period 1 utility of a lottery π with payoffs in period 2. Then V is given by:

$$
(8) \qquad V(\pi) = \begin{cases} \mathbb{E}[h(v(z,\cdot))] & \text{for early uncertainty resolution} \\[2mm] h(\mathbb{E}[v(z,\cdot)]) & \text{for late uncertainty resolution} \end{cases}
$$

Note that the expectations operator is always applied to the quantity that is known at the end of period one. If uncertainty resolves early, the decision-maker applies the weighting

function to the utility of the specific outcomes of π. If the outcome of π remains uncertain until the second period, he applies the weighting function to its expected value. Kreps and Porteus (1978) show that h is strictly convex (concave) if and only if the decision maker always prefers early to late (late to early) resolution, and is linear if the decision maker is indifferent.

We use a power specification for $h(\cdot)$ and a slightly modified version of (2) for period 2 utility. These modifications become necessry to handle risk aversion and risk lovingness, and negative payoffs within the Kreps-Porteus framework. The curvature of $h(\cdot)$ is governed by a parameter ρ, which thus captures preferences towards uncertainty resolution. In particular, we use:

$$(9) \qquad v(z, \gamma, \lambda, \rho) \;=\; \begin{cases} \max\{-\frac{\lambda}{\gamma}, 0\} - \frac{1}{\gamma} e^{-\gamma \rho^S z} & \text{for } z \geq 0 \\[2mm] \max\{-\frac{\lambda}{\gamma}, 0\} + \frac{\lambda-1}{\gamma} - \frac{\lambda}{\gamma} e^{-\gamma \rho^S z} & \text{for } z < 0 \end{cases}$$

where S is the sign operator given by:

$$(10) \qquad S \;=\; \begin{cases} 1 & \text{for } \gamma \geq 0 \\ -1 & \text{for } \gamma < 0. \end{cases}$$

We choose the following parsimonious "power function" specification of the weighting function h:

$$(11) \qquad h(v(z, \cdot)) \;=\; -S(-S\, v(z, \cdot))^{\rho^{-S}}$$

where $\rho \in \mathbb{R}_+$. For $\rho > 1$, $h(\cdot)$ is convex and early resolution is preferred to late resolution. Indifference is obtained for $\rho = 1$, and late resolution is preferred for $\rho < 1$.

The building blocks of (9) and (11) seem complicated because of the necessity to accommodate both types of utility curvature. The term $-\frac{\lambda}{\gamma}$ is added for risk lovers to assure that the weighting function $h(\cdot)$ can be applied, i.e. it guarantees that $v(z, \gamma, \lambda, \rho)$ is always greater than zero for $\gamma < 0$. Including ρ^S in the exponent serves to retain the interpretation of $\gamma > 0$ as the coefficient of absolute risk aversion for early resolving lotteries on the positive domain. For such lotteries $V(\pi)$ collapses to $\mathbb{E}[u(\pi)]$ given in (2) if the subject is not risk loving. This implies that the distinction between risk aversion and uncertainty resolution timing preferences is identified for risk averse subjects if there are gambles on the positive domain and the timing of uncertainty resolution is varied. Note that this distinction is only approximately true for gambles with negative outcomes because of the additive term $\frac{\lambda-1}{\gamma}$ in (9). For risk lovers, the inclusion of $-\frac{\lambda}{\gamma}$ distorts the interpretation of γ by the same token. For the parameter values that we estimate, the magnitudes of the distortions are small. The Köbberling and Wakker (2005) definition of loss aversion remains valid for period 2 utility.

7 Alternative utility specifications: Results

The parameter estimates corresponding to Table 4 for the alternative models are presented in Tables 7–9. These estimates indicate that the results based on the preferred loss aversion specification are robust to changing the functional form. Across all specifications, the parameter estimates of γ, λ and ω are significantly lower for the lab sample than for the Internet

sample. Moreover, the results for the Internet-uni sample mirror those of the lab sample, verifying that the behavioral differences between the lab and the Internet samples are mainly driven by the different composition of the samples with respect to age and education.

The effects of self-selection that we observed for the loss aversion specification are also robust to changing the set of covariates in the probit selection regression. Table 18 displays the significance of the selection weights using the narrow set of covariates. for the alternative models. Overall, the effects are similar to those observed for the loss aversion specification reported in Table 6. Again, the strongest self-selection effect relates to the tendency to make errors. The effects on the other parameters are smaller and less significant. Regarding the source of the selection bias, the estimates using the small number of covariates confirm that most of the selection effect is driven by steps 2 and 3, while step 1 contributes to a lesser extent.

8   Additional robustness checks

In this section we briefly comments on the results of some additional robustness checks presented in the appendix. Table 12 presents the estimates of the preferred loss aversion model using a more extensive set of covariates. The lab dummy is now considerably smaller, and insignificant, giving another indication that the socio-demographic composition is the main driver behind the Lab vs Internet discrepancy. Controlling for age and education, the observed difference between the laboratory and the Internet samples vanishes.

Since there was a small difference with respect to show-up fees between the lab and the Internet experiments in hypothetical treatment, we chose to present results based on only the incentivised treatments in the main text. Results based on the full sample, including the hypothetical treatment, are presented in Table 13. With respect to selection effects, the results from the full sample are practically identical to those of the restricted sample previously considered in Table 4. Also the estimates using an unrestricted covariance matrix presented in Tables 10 and 11 are very similar to the ones obtained by using the diagonal restriction.

The general conclusion regarding the effects of self selection also appears to be robust towards the choice of empirical strategy. As mentioned in the text, if we follow Harrison *et al.* (2009) and take the approach developed by Heckman (1976, 1979), the general result echoes that based on the MAR assumption. In Tables 20-25, the estimates of the Heckman selection model are presented for each parameter, conditioning on both the broad and a narrow set of covariates used previously. Acknowledging the potential relation between region of residence and our preference parameter, we nevertheless follow Harrison *et al.* (2009) for comparison purposes and use regional dummies as exclusion restrictions. Individual-level parameters are obtained by taking the mean parameter estimate from the structural model, conditioning on the actual choices of each individual. Regardless of whether we consider the broad or narrow set of covariates in Table 5, the selection effect is only significant for the error parameter.

In the model displayed in Table 4, the two laboratory samples were merged into one motivated by the observation that there is no significant difference between the two lab treatments. This result is demonstrated by the estimation results distinguishing between the Lab-Lab and the Lab-Internet treatment displayed in Table 14. The coefficients of the dummy variable for the Lab-Lab treatment are insignificant, indicating that the environmental

treatment variables had no effect on observed behavior.

9 Further documentation of the experiments

Figure 6: Translations of the Welcome Screens in the CentERpanel Experiment

High (Low) Incentive Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

You will receive 15 (5) Euros for participating. Then you can, depending on the choices you make and on chance, earn more or lose part of the 15 (5) Euros. If completing the total experiment, you receive the reward for participating, possibly increased by your gain (or reduced by your loss) in one of the choices you have made. Whether the latter occurs and which choice then determines your payoff, will be determined by chance. **Your total reward will be added to your CentERpoints.**

The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences. Assume in each choice problem that this choice determines your actual payoff.

This questionnaire is about making choices, and your payoff depends on your choices and on chance. If you do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

O No, I do **not** want to complete this questionnaire

Hypothetical Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences.

This questionnaire is about making choices between several situations in which you can (hypothetically) gain or lose money. Your revenue depends on the choices you make and on chance. **What matters is what you would do in hypothetical situations, in reality, there is nothing at stake for you.** If you nevertheless do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

O No, I do **not** want to complete this questionnaire

10 Figures corresponding to Figure 3 with upper and lower switchpoints

Figure 7: Mean (lowest) switch points in the CentERpanel and laboratory experiments
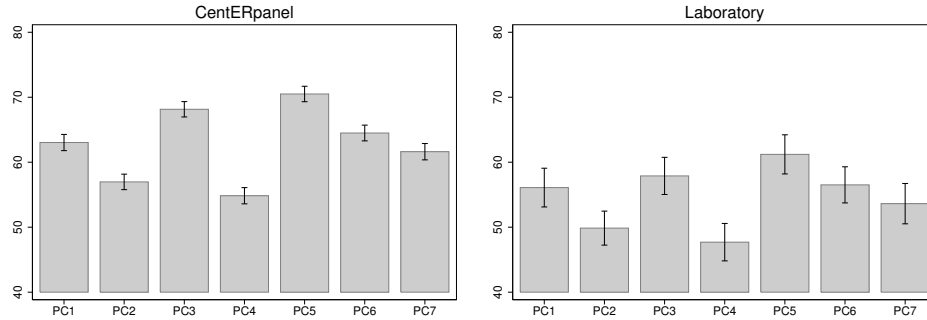
Note: The numbering of the payoff configurations (PC) conforms to those in Table 1. "Switch points" are defined as highest probability corresponding to an 'A' choice that is still lower than the minimum probability with a choice of 'B'. Error bars depict 95% confidence intervals.

Figure 8: Mean (highest) switch points in the CentERpanel and laboratory experiments
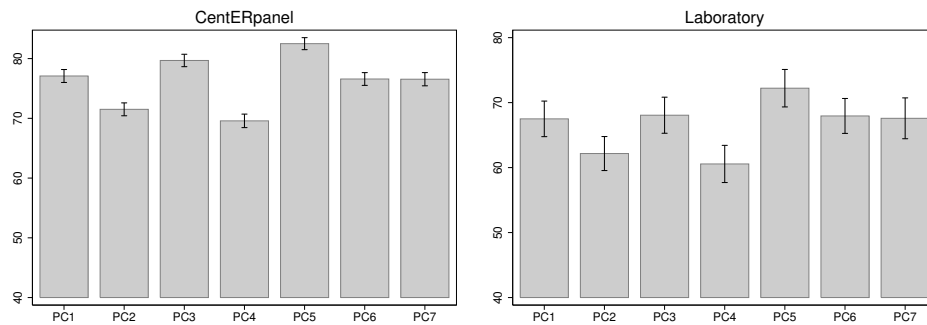
Note: The numbering of the payoff configurations (PC) conforms to those in Table 1. "Switch points" are defined as the minimum probability with a 'B' choice that is still higher than the maximum probability where option 'A' was chosen. Error bars depict 95% confidence intervals.

11 Alternative functional forms

Table 7: Exponential utility specification (1), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	ω
Constant	0.0471***	0.0965***
	(0.0016)	(0.0102)
Internet university subsample	-0.0166**	-0.0712***
	(0.0072)	(0.0109)
Participant in Lab experiment	-0.0234***	-0.0860***
	(0.0071)	(0.0092)
Low incentive treatment†	2.67***	1.09
	(0.0941)	(0.140)
σ	0.0451***	2.04***
	(0.0013)	(0.120)

Note The value of the log-likelihood function is 29410.6, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 4.65 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

† The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .29 (.01).

Table 8: Prospect theory utility specification (7), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0355***	3.12***	0.0824***
	(0.0012)	(0.153)	(0.0085)
Internet university subsample	-0.0194***	-0.702	-0.0578***
	(0.0040)	(0.448)	(0.0099)
Participant in Lab experiment	-0.0174***	-1.16***	-0.0691***
	(0.0048)	(0.365)	(0.0077)
Low incentive treatment[†]	2.54***	1.06	1.21
	(0.0738)	(0.0399)	(0.150)
σ	0.0392***	1.22***	2.04***
	(0.0008)	(0.0413)	(0.110)

Note The value of the log-likelihood function is 28172.3, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 4.05 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

Table 9: Kreps-Porteus utility specification, (9) and (11), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ρ	ω
Constant	0.0325***	2.36***	1.01	0.0899***
	(0.0008)	(0.109)	(0.0218)	(0.0081)
Internet university subsample	-0.0073**	0.439	0.115	-0.0671***
	(0.0031)	(0.449)	(0.0914)	(0.0094)
Participant in Lab experiment	-0.0138***	-0.591*	0.0647	-0.0743***
	(0.0024)	(0.306)	(0.0857)	(0.0076)
Low incentive treatment[†]	2.94***	1.04	1.000	1.08
	(0.0806)	(0.0341)	(0.0012)	(0.125)
σ	0.0348***	1.30***	0.459***	1.97***
	(0.0008)	(0.0429)	(0.0267)	(0.103)

Note The value of the log-likelihood function is 27991.4, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 3.82 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

12 Unrestricted covariance matrix

Table 10: Loss aversion utility specification (2), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, unrestricted covariance matrix

Covariate	γ	λ	ω
Constant	0.0353***	2.73***	0.0815***
	(0.0011)	(0.143)	(0.0083)
Internet university subsample	-0.0178***	-0.586*	-0.0652***
	(0.0042)	(0.339)	(0.0086)
Participant in Lab experiment	-0.0188***	-1.12***	-0.0687***
	(0.0042)	(0.307)	(0.0076)
Low incentive treatment[†]	2.65***	1.04	1.23
	(0.0766)	(0.0383)	(0.151)
σ	0.0364***	1.28***	2.11***
	(0.0009)	(0.0459)	(0.116)

Note The value of the log-likelihood function is 28143.7, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 4.18 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

Table 11: Correlation matrix for: Loss aversion utility specification (2), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, unrestricted covariance matrix

	σ_γ	σ_λ	σ_ω
σ_γ	0.036	0.136	0.068
	(0.001)	(0.036)	(0.029)
σ_λ		1.283	0.291
		(0.046)	(0.044)
σ_ω			2.113
			(0.116)
τ	4.176		
	(0.072)		
$\tau_{\text{Low Inc}}$ [†]	0.282		
	(0.007)		
Log likelihood	28143.7		

13   Extended set of regressors

Table 12: Loss aversion utility specification (2), CentERpanel and laboratory experiments combined, only real incentive treatments, full set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0160***	1.35	0.0264***
	(0.0052)	(0.296)	(0.0093)
Female	0.0115***	0.362**	0.0076*
	(0.0020)	(0.141)	(0.0044)
Primary / lower secondary education	0.0170***	0.351	0.0467***
	(0.0041)	(0.240)	(0.0129)
Higher sec. educ. / interm. voc. training	0.0109***	0.417*	0.0070
	(0.0040)	(0.216)	(0.0065)
Higher vocational training	0.0068*	0.320	-0.0002
	(0.0040)	(0.229)	(0.0065)
Age 35-44 years	-0.0028	0.448**	0.0129*
	(0.0031)	(0.216)	(0.0074)
Age 45-54 years	0.0039	-0.0098	0.0143*
	(0.0029)	(0.180)	(0.0078)
Age 55-64 years	0.0030	0.190	0.0343***
	(0.0032)	(0.217)	(0.0122)
Age 65 years and older	0.0079**	-0.0014	0.115***
	(0.0040)	(0.238)	(0.0269)
Working	0.0014	0.253	0.0000
	(0.0025)	(0.161)	(0.0043)
Unemployed and looking for a job	0.0100	0.677	0.0051
	(0.0079)	(0.501)	(0.0151)
Internet university subsample	-0.0123**	0.196	-0.0083
	(0.0052)	(0.333)	(0.0081)
Participant in Lab experiment	-0.0005	0.0671	-0.0155*
	(0.0065)	(0.379)	(0.0090)
Took < 9 min. for experiment			0.0255**
			(0.0103)
Took > 18 min. for experiment			-0.0135***
			(0.0052)
Low incentive treatment†	2.65***	1.01	1.35
	(0.0851)	(0.0175)	(0.272)
σ	0.0382***	1.22***	1.90***
	(0.0009)	(0.0419)	(0.105)

Note The value of the log-likelihood function is 28070.5, based on 1345 individuals who made 69548 choices. The value of τ is estimated to be 4.19 (.07). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

† The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

14   Estimations based on the full sample

Table 13: Loss aversion utility specification (2), CentERpanel and laboratory experiments combined, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0326***	2.34***	0.0861***
	(0.0010)	(0.120)	(0.0074)
Internet university subsample	-0.0124***	0.150	-0.0659***
	(0.0029)	(0.328)	(0.0078)
Participant in Lab experiment	-0.0155***	-0.891***	-0.0699***
	(0.0036)	(0.227)	(0.0065)
Hypothetical treatment	0.0002	0.980***	0.0190*
	(0.0015)	(0.240)	(0.0106)
Low incentive treatment[†]	2.87***	0.993	1.13
	(0.0754)	(0.0270)	(0.125)
σ	0.0367***	1.32***	1.99***
	(0.0007)	(0.0313)	(0.0802)

Note The value of the log-likelihood function is 41620, based on 1965 individuals who made 101528 choices. The value of τ is estimated to be 4.19 (.05). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .28 (.01).

15 Estimations distinguishing between the "lab-lab" and the "lab-internet" treatments

Table 14: Loss aversion utility specification (2), Laboratory experiment, all incentive treatments, dummies for implementation mode treatment, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0152***	1.83**	0.0043
	(0.0034)	(0.325)	(0.0046)
Lab-lab treatment	0.0043	0.286	0.0007
	(0.0036)	(0.386)	(0.0033)
Hypothetical treatment	0.0015	-0.0434	0.0042
	(0.0040)	(0.382)	(0.0055)
Low incentive treatment†	2.79***	0.812**	1.14
	(0.244)	(0.0745)	(2.02)
σ	0.0204***	0.933***	2.65***
	(0.0012)	(0.0809)	(0.779)

Note The value of the log-likelihood function is 3546.2, based on 178 individuals who made 9820 choices. The value of τ is estimated to be 4.44 (.13). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

† The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .29 (.02).

16 Separate estimations for CentERpanel and Laboratory samples

Table 15: Loss aversion utility specification (2), CentERpanel experiment, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0337***	2.27***	0.0763***
	(0.0012)	(0.112)	(0.0085)
Low incentive treatment[†]	2.58***	0.950*	1.15
	(0.0918)	(0.0284)	(0.162)
σ	0.0424***	1.27***	2.05***
	(0.0012)	(0.0456)	(0.117)

Note The value of the log-likelihood function is 25967.5, based on 1232 individuals who made 63292 choices. The value of τ is estimated to be 4.13 (.08). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .29 (.01).

Table 16: Loss aversion utility specification (2), Laboratory experiment, only real incentive treatments, minimal set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0162***	1.73***	0.0071
	(0.0025)	(0.191)	(0.0091)
Low incentive treatment[†]	2.77***	0.947	0.702
	(0.271)	(0.0710)	(1.26)
σ	0.0217***	0.711***	2.49**
	(0.0017)	(0.0945)	(1.05)

Note The value of the log-likelihood function is 2193.6, based on 113 individuals who made 6256 choices. The value of τ is estimated to be 4.29 (.19). Standard errors are in parentheses.

Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.

[†] The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .3 (.02).

17 Marginal effects in the selection model

Table 17: Self-selection into the CentERpanel experiment: Marginal effects

Specification	NP (1)	DO/SP (2)	NP (3)	DO/SP (4)
Low incentive treatment	-.020	-.041	-.026	-.041
High incentive treatment	-.028	-.031	-.027	-.034
Female	.033	.024	.022	.025
Primary / lower secondary education	.053	.066	.069	.050
Higher sec. educ. / interm. voc. training	.039	.040	.024	.040
Higher vocational training	-.031	.038	-.022	.034
Age 35-44 years	-.052	.029	-.048	.031
Age 45-54 years	-.098	.058	-.082	.067
Age 55-64 years	-.109	.050	-.090	.072
Age 65 years and older	-.107	.133	-.098	.162
Working	-.011	-.020	.001	-.003
Unemployed and looking for a job	.025	.002	.011	.007
Net household income ∈ [22k Euros; 40k Euros)			.014	.009
Net household income at least 40k Euros			.063	-.006
Assets worth ∈ [10k Euros, 50k Euros)			.054	-.013
Assets worth ∈ [50k Euros, 200k Euros)			.006	.013
Assets worth more than 200k Euros			.041	.010
Household financial administrator			-.027	-.016
Employer offers savings plan			-.052	.011
Has sav. plan via employer			.018	-.035
Has sav. acc. or similar			.019	.009
Holds stocks, or similar			.004	-.020
Constant	.152	.066	.126	.063
Number of observations	2295	2295	1690	1690

Note: Marginal effects corresponding to multinomial logit regression from Table 5. Columns indicate categories of the dependent variable by regression type. Columns (1) and (3) list estimates for opting for non-participation on the first screen (NP); columns (2) and (4) those for dropping out before completion (DO) or finishing the experiment in less than 5:20 minutes (SP).

18 Selection weights with reduced set of covariates

Table 18: Significance of selection weights, loss aversion utility specification (2), diagonal covariance matrix,

Parameter / Weight	Average	P-value, steps 2, 3	P-value, no weight
γ – Weights for steps 1 to 3	0.0522	0.516	0.376
	(0.0694)		
γ – Weights for steps 2 and 3	0.0517	.	0.036
	(0.0703)		
γ – Unweighted	0.0532	.	.
	(0.0742)		
λ – Weights for steps 1 to 3	11.35	0.575	0.007
	(28.68)		
λ – Weights for steps 2 and 3	11.21	.	0.001
	(29.06)		
λ – Unweighted	10.45	.	.
	(26.64)		
ω – Weights for steps 1 to 3	0.238	0.022	0.000
	(0.250)		
ω – Weights for steps 2 and 3	0.231	.	0.000
	(0.246)		
ω – Unweighted	0.217	.	.
	(0.240)		

Note: Treatment effects are netted out. Standard errors are given in parenthesis. P-values are based on t-tests, described in Footnote 15 in Section 4.2, with the null hypotheses of equal means.

19 Coefficient estimates for selection specification used in Section 4.2

Table 19: Loss aversion utility specification (2), CentERpanel experiment, self-selection specification, full set of covariates, diagonal covariance matrix

Covariate	γ	λ	ω
Constant	0.0104*	2.44*	0.0242**
	(0.0059)	(0.830)	(0.0101)
Female	0.0044**	0.275	0.0051
	(0.0021)	(0.286)	(0.0037)
Primary / lower secondary education	0.0181***	-0.0849	0.0447***
	(0.0037)	(0.440)	(0.0160)
Higher sec. educ. / interm. voc. training	0.0145***	0.0635	0.0161**
	(0.0036)	(0.412)	(0.0077)
Higher vocational training	0.0107***	-0.173	0.0082
	(0.0037)	(0.405)	(0.0065)
Age 35-44 years	0.0036	0.795	0.0115
	(0.0030)	(0.503)	(0.0072)
Age 45-54 years	0.0067**	-0.408	0.0214**
	(0.0032)	(0.377)	(0.0107)
Age 55-64 years	0.0018	0.129	0.0515**
	(0.0034)	(0.461)	(0.0209)
Age 65 years and older	0.0146***	-0.612	0.115***
	(0.0041)	(0.477)	(0.0404)
Working	-0.0016	0.643	0.0041
	(0.0025)	(0.438)	(0.0046)
Unemployed and looking for a job	0.0188***	0.464	-0.0046
	(0.0070)	(1.06)	(0.0099)
Net household income $\in$ [22k Euros; 40k Euros)	-0.0011	-0.351	-0.0053
	(0.0023)	(0.308)	(0.0039)
Net household income at least 40k Euros	-0.0040	-0.932**	-0.0086*
	(0.0032)	(0.441)	(0.0052)
Assets worth $\in$ [10k Euros, 50k Euros)	0.0057*	0.249	-0.0078*
	(0.0033)	(0.449)	(0.0046)
Assets worth $\in$ [50k Euros, 200k Euros)	-0.0037	-0.217	-0.0033
	(0.0027)	(0.320)	(0.0036)
Assets worth more than 200k Euros	0.0017	0.0170	-0.0045
	(0.0032)	(0.422)	(0.0040)
Household financial administrator	0.0005	-0.250	-0.0034
	(0.0021)	(0.287)	(0.0034)
Employer offers savings plan	-0.0039	0.211	0.0026
	(0.0034)	(0.487)	(0.0063)
Has sav. plan via employer	0.0109***	-0.560	-0.0037
	(0.0034)	(0.402)	(0.0051)
Has sav. acc. or similar	0.0031	0.238	0.0105
	(0.0035)	(0.594)	(0.0069)
Holds stocks, or similar	-0.0076***	-0.0018	-0.0051
	(0.0021)	(0.291)	(0.0034)
Took < 9 min. for experiment			0.0256**
			(0.0109)
Took > 18 min. for experiment			-0.0119**
			(0.0053)
Hypothetical treatment	0.0015	1.19**	0.0036
	(0.0020)	(0.552)	(0.0038)
Low incentive treatment†	2.84***	0.903*	1.18
	(0.0969)	(0.0509)	(0.269)
σ	0.0369***	1.45***	1.83***
	(0.0009)	(0.0510)	(0.0901)

Note The value of the log-likelihood function is 28503.9, based on 1346 individuals who made 69124 choices. The value of τ is estimated to be 4.22 (.07). Standard errors are in parentheses.
Regression coefficients are transformed back to the original scale. In other words, the constant is defined by $g_\eta(\beta_1^\eta)$ and represents median parameters in the left-out categories. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories: $g_\eta(\beta_{\text{constant}}^\eta + \beta_{\text{dummy}}^\eta) - g_\eta(\beta_{\text{constant}}^\eta)$.
† The low incentive treatment enters multiplicatively and we report the non-transformed coefficients, i.e. $\beta_{\text{low incentive}}^\lambda$ Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. For τ, this parameter is estimated to be .3 (.01).

20 Results from a Heckman selection model

Table 20: Sample selection model, main equation, γ-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	0.050***	(0.006)	0.047***	(0.007)
High incentive treatment	-0.003	(0.006)	-0.003	(0.007)
Female	0.017***	(0.004)	0.016***	(0.005)
Primary / lower secondary education	0.024***	(0.008)	0.025***	(0.008)
Higher sec. educ. / interm. voc. training	0.018***	(0.007)	0.018**	(0.007)
Higher vocational training	0.014**	(0.006)	0.014**	(0.007)
Age 35-44 years	-0.000	(0.005)	0.000	(0.006)
Age 45-54 years	0.005	(0.005)	0.007	(0.006)
Age 55-64 years	0.005	(0.006)	0.008	(0.007)
Age 65 years and older	0.007	(0.006)	0.012	(0.008)
Working	-0.003	(0.005)	-0.003	(0.005)
Unemployed and looking for a job	0.014	(0.011)	0.019	(0.013)
Net household income $\in [22\text{k Euros}; 40\text{k Euros})$			-0.001	(0.005)
Net household income at least 40k Euros			-0.000	(0.006)
Assets worth $\in [10\text{k Euros}, 50\text{k Euros})$			0.006	(0.006)
Assets worth $\in [50\text{k Euros}, 200\text{k Euros})$			-0.000	(0.005)
Assets worth more than 200k Euros			-0.003	(0.006)
Household financial administrator			0.002	(0.005)
Employer offers savings plan			-0.002	(0.008)
Has sav. plan via employer			0.009	(0.008)
Has sav. acc. or similar			0.006	(0.008)
Holds stocks, or similar			-0.009*	(0.005)
Constant	0.016	(0.015)	0.011	(0.017)
Inverse Mills ratio	-0.018	(0.031)	-0.024	(0.032)
Observations	2194		1690	

Note: Coefficient estimates of the main equation of the Heckman two-step sample selection model. Dependent variable is the median γ-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level.

Table 21: Sample selection model, selection equation, γ-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	0.390***	(0.073)	0.449***	(0.086)
High incentive treatment	0.331***	(0.076)	0.366***	(0.090)
Female	-0.200***	(0.063)	-0.180**	(0.075)
Primary / lower secondary education	-0.372***	(0.118)	-0.356**	(0.146)
Higher sec. educ. / interm. voc. training	-0.243**	(0.117)	-0.207	(0.143)
Higher vocational training	-0.086	(0.123)	-0.106	(0.147)
Age 35-44 years	0.106	(0.096)	0.092	(0.120)
Age 45-54 years	0.169*	(0.092)	0.112	(0.115)
Age 55-64 years	0.211**	(0.100)	0.140	(0.122)
Age 65 years and older	-0.113	(0.106)	-0.137	(0.132)
Working	0.122	(0.079)	0.004	(0.100)
Unemployed and looking for a job	0.069	(0.204)	-0.059	(0.239)
Province==Friesland	0.438**	(0.211)	0.498**	(0.251)
Province==Drenthe	0.102	(0.220)	0.103	(0.264)
Province==Overijssel	0.233	(0.190)	0.405*	(0.232)
Province==Flevoland	0.223	(0.254)	0.217	(0.314)
Province==Gelderland	0.204	(0.176)	0.185	(0.213)
Province==Utrecht	0.470**	(0.212)	0.519**	(0.258)
Province==Noord-Holland	0.482***	(0.177)	0.376*	(0.210)
Province==Zuid-Holland	0.338**	(0.170)	0.314	(0.203)
Province==Zeeland	0.181	(0.221)	0.047	(0.262)
Province==Noord-Brabant	0.196	(0.172)	0.112	(0.205)
Province==Limburg	0.330*	(0.189)	0.378*	(0.227)
Net household income $\in$ [22k Euros; 40k Euros)			-0.083	(0.086)
Net household income at least 40k Euros			-0.139	(0.121)
Assets worth $\in$ [10k Euros, 50k Euros)			-0.051	(0.118)
Assets worth $\in$ [50k Euros, 200k Euros)			-0.089	(0.097)
Assets worth more than 200k Euros			-0.102	(0.114)
Household financial administrator			0.196**	(0.080)
Employer offers savings plan			0.102	(0.143)
Has sav. plan via employer			0.246*	(0.148)
Has sav. acc. or similar			-0.057	(0.153)
Holds stocks, or similar			0.127	(0.083)
Constant	0.456**	(0.201)	0.476	(0.291)
Observations	2194		1690	

Note: Coefficient estimates of the selection equation of the Heckman two-step sample selection model. Dependent variable is the median γ-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level. Left out province dummy is Groningen.

Table 22: Sample selection model, main equation, λ-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	-3.544	(2.604)	-3.463	(2.989)
High incentive treatment	-1.947	(2.432)	-2.638	(2.817)
Female	1.321	(1.747)	1.475	(1.912)
Primary / lower secondary education	-0.772	(3.077)	-0.814	(3.521)
Higher sec. educ. / interm. voc. training	-2.457	(2.648)	-3.446	(3.081)
Higher vocational training	0.575	(2.455)	-1.073	(2.949)
Age 35-44 years	7.143***	(2.163)	8.290***	(2.632)
Age 45-54 years	-0.844	(2.216)	-1.835	(2.591)
Age 55-64 years	0.063	(2.479)	-0.167	(2.855)
Age 65 years and older	-4.599*	(2.562)	-5.529*	(3.269)
Working	-0.593	(1.895)	-1.010	(2.262)
Unemployed and looking for a job	-2.371	(4.484)	-3.332	(5.349)
Net household income $\in [22\text{k Euros}; 40\text{k Euros})$			-1.976	(1.981)
Net household income at least 40k Euros			-4.565*	(2.723)
Assets worth $\in [10\text{k Euros}, 50\text{k Euros})$			-0.449	(2.625)
Assets worth $\in [50\text{k Euros}, 200\text{k Euros})$			-0.742	(2.246)
Assets worth more than 200k Euros			-1.705	(2.675)
Household financial administrator			3.710*	(2.126)
Employer offers savings plan			2.489	(3.256)
Has sav. plan via employer			-2.273	(3.376)
Has sav. acc. or similar			3.525	(3.180)
Holds stocks, or similar			0.177	(1.969)
Constant	6.407	(5.868)	4.828	(7.350)
Inverse Mills ratio	17.243	(12.340)	17.489	(13.367)
Observations	2194		1690	

Note: Coefficient estimates of the main equation of the Heckman two-step sample selection model. Dependent variable is the median λ-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level.

Table 23: Sample selection model, selection equation, λ-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	0.390***	(0.073)	0.449***	(0.086)
High incentive treatment	0.331***	(0.076)	0.366***	(0.090)
Female	-0.200***	(0.063)	-0.180**	(0.075)
Primary / lower secondary education	-0.372***	(0.118)	-0.356**	(0.146)
Higher sec. educ. / interm. voc. training	-0.243**	(0.117)	-0.207	(0.143)
Higher vocational training	-0.086	(0.123)	-0.106	(0.147)
Age 35-44 years	0.106	(0.096)	0.092	(0.120)
Age 45-54 years	0.169*	(0.092)	0.112	(0.115)
Age 55-64 years	0.211**	(0.100)	0.140	(0.122)
Age 65 years and older	-0.113	(0.106)	-0.137	(0.132)
Working	0.122	(0.079)	0.004	(0.100)
Unemployed and looking for a job	0.069	(0.204)	-0.059	(0.239)
Province==Friesland	0.438**	(0.211)	0.498**	(0.251)
Province==Drenthe	0.102	(0.220)	0.103	(0.264)
Province==Overijssel	0.233	(0.190)	0.405*	(0.232)
Province==Flevoland	0.223	(0.254)	0.217	(0.314)
Province==Gelderland	0.204	(0.176)	0.185	(0.213)
Province==Utrecht	0.470**	(0.212)	0.519**	(0.258)
Province==Noord-Holland	0.482***	(0.177)	0.376*	(0.210)
Province==Zuid-Holland	0.338**	(0.170)	0.314	(0.203)
Province==Zeeland	0.181	(0.221)	0.047	(0.262)
Province==Noord-Brabant	0.196	(0.172)	0.112	(0.205)
Province==Limburg	0.330*	(0.189)	0.378*	(0.227)
Net household income $\in$ [22k Euros; 40k Euros)			-0.083	(0.086)
Net household income at least 40k Euros			-0.139	(0.121)
Assets worth $\in$ [10k Euros, 50k Euros)			-0.051	(0.118)
Assets worth $\in$ [50k Euros, 200k Euros)			-0.089	(0.097)
Assets worth more than 200k Euros			-0.102	(0.114)
Household financial administrator			0.196**	(0.080)
Employer offers savings plan			0.102	(0.143)
Has sav. plan via employer			0.246*	(0.148)
Has sav. acc. or similar			-0.057	(0.153)
Holds stocks, or similar			0.127	(0.083)
Constant	0.456**	(0.201)	0.476	(0.291)
Observations	2194		1690	

Note: Coefficient estimates of the selection equation of the Heckman two-step sample selection model. Dependent variable is the median λ-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level. Left out province dummy is Groningen.

Table 24: Sample selection model, main equation, ω-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	0.002	(0.023)	0.008	(0.025)
High incentive treatment	0.014	(0.021)	0.013	(0.023)
Female	0.009	(0.015)	-0.005	(0.016)
Primary / lower secondary education	0.098***	(0.027)	0.071**	(0.029)
Higher sec. educ. / interm. voc. training	0.031	(0.023)	0.010	(0.026)
Higher vocational training	0.022	(0.022)	0.011	(0.025)
Age 35-44 years	0.032*	(0.019)	0.043*	(0.022)
Age 45-54 years	0.047**	(0.020)	0.059***	(0.022)
Age 55-64 years	0.090***	(0.022)	0.098***	(0.024)
Age 65 years and older	0.114***	(0.023)	0.120***	(0.027)
Working	0.014	(0.017)	0.014	(0.019)
Unemployed and looking for a job	-0.020	(0.040)	-0.051	(0.045)
Net household income $\in [22\text{k Euros}; 40\text{k Euros})$			-0.032*	(0.016)
Net household income at least 40k Euros			-0.042*	(0.023)
Assets worth $\in [10\text{k Euros}, 50\text{k Euros})$			-0.046**	(0.022)
Assets worth $\in [50\text{k Euros}, 200\text{k Euros})$			-0.038**	(0.019)
Assets worth more than 200k Euros			-0.049**	(0.022)
Household financial administrator			-0.000	(0.018)
Employer offers savings plan			0.017	(0.027)
Has sav. plan via employer			-0.009	(0.028)
Has sav. acc. or similar			0.017	(0.027)
Holds stocks, or similar			-0.019	(0.016)
Constant	0.040	(0.052)	0.093	(0.061)
Inverse Mills ratio	0.175	(0.109)	0.182	(0.111)
Observations	2194		1690	

Note: Coefficient estimates of the main equation of the Heckman two-step sample selection model. Dependent variable is the median ω-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level.

Table 25: Sample selection model, selection equation, ω-estimates based on the loss aversion utility specification (2).

	(1)		(2)	
Low incentive treatment	0.390***	(0.073)	0.449***	(0.086)
High incentive treatment	0.331***	(0.076)	0.366***	(0.090)
Female	-0.200***	(0.063)	-0.180**	(0.075)
Primary / lower secondary education	-0.372***	(0.118)	-0.356**	(0.146)
Higher sec. educ. / interm. voc. training	-0.243**	(0.117)	-0.207	(0.143)
Higher vocational training	-0.086	(0.123)	-0.106	(0.147)
Age 35-44 years	0.106	(0.096)	0.092	(0.120)
Age 45-54 years	0.169*	(0.092)	0.112	(0.115)
Age 55-64 years	0.211**	(0.100)	0.140	(0.122)
Age 65 years and older	-0.113	(0.106)	-0.137	(0.132)
Working	0.122	(0.079)	0.004	(0.100)
Unemployed and looking for a job	0.069	(0.204)	-0.059	(0.239)
Province==Friesland	0.438**	(0.211)	0.498**	(0.251)
Province==Drenthe	0.102	(0.220)	0.103	(0.264)
Province==Overijssel	0.233	(0.190)	0.405*	(0.232)
Province==Flevoland	0.223	(0.254)	0.217	(0.314)
Province==Gelderland	0.204	(0.176)	0.185	(0.213)
Province==Utrecht	0.470**	(0.212)	0.519**	(0.258)
Province==Noord-Holland	0.482***	(0.177)	0.376*	(0.210)
Province==Zuid-Holland	0.338**	(0.170)	0.314	(0.203)
Province==Zeeland	0.181	(0.221)	0.047	(0.262)
Province==Noord-Brabant	0.196	(0.172)	0.112	(0.205)
Province==Limburg	0.330*	(0.189)	0.378*	(0.227)
Net household income $\in$ [22k Euros; 40k Euros)			-0.083	(0.086)
Net household income at least 40k Euros			-0.139	(0.121)
Assets worth $\in$ [10k Euros, 50k Euros)			-0.051	(0.118)
Assets worth $\in$ [50k Euros, 200k Euros)			-0.089	(0.097)
Assets worth more than 200k Euros			-0.102	(0.114)
Household financial administrator			0.196**	(0.080)
Employer offers savings plan			0.102	(0.143)
Has sav. plan via employer			0.246*	(0.148)
Has sav. acc. or similar			-0.057	(0.153)
Holds stocks, or similar			0.127	(0.083)
Constant	0.456**	(0.201)	0.476	(0.291)
Observations	2194		1690	

Note: Coefficient estimates of the selection equation of the Heckman two-step sample selection model. Dependent variable is the median ω-estimate based on the loss aversion utility specification (2), full sample, diagonal covariance matrix. Standard errors are given in parenthesis. The final selected sample consists of invited participants who decide to participate and take more than 5:20 minutes to complete the experiment. Asterisks indicate significance at the 10%, 5%, and 1%-level. Left out province dummy is Groningen.