The standard portfolio choice problem in Germany*

Christoph Breunig† Steffen Huck‡ Tobias Schmidt§ Georg Weizsäcker¶

Emory University | WZB Berlin and UCL | QuantCo | Humboldt-Universität zu Berlin

November 19, 2020

Abstract

We study an investment experiment with a representative sample of German households. Respondents invest in a safe asset and a risky asset whose return is tied to the German stock market. Experimental investments correlate with beliefs about stock market returns and exhibit desirable external validity at least in one respect: they predict real-life stock market participation. But many households are unresponsive to an exogenous increase in the risky asset's return. The data analysis and a series of additional laboratory experiments suggest that task complexity decreases the responsiveness to incentives. Modifying the safe asset's return has a larger effect on behaviour than modifying the risky asset's return.

*Tanika Chakraborty was involved in an early stage of this research and we are grateful for her many valuable contributions to it. We thank audiences at Berkeley, WZB Berlin, Bonn, CESifo, LMU Munich, Stanford and the University of Zurich as well as Justin Valasek and Muriel Niederle for helpful comments. We also thank Jürgen Schupp, David Richter, Elisabeth Liebau, Nico Siegel and the staff of TNS Infratest for their help and expertise in preparing and administering the SOEP-IS survey module and colleagues at the WZB-TU decision laboratory in Berlin for their excellent contributions in the preparation and conduct of the experiments. Ethical approval was obtained from the DIW Berlin. Financial support by the ERC (Starting Grant 263412) and the German Science Foundation (Collaborative Research Center TRR 190, "Rationality and Competition") is gratefully acknowledged.

†Department of Economics, Emory University, Rich Memorial Building, Atlanta, GA 30322, USA, christoph.breunig@emory.edu

‡WZB Berlin, Reichpietschufer 50, 10785 Berlin, steffen.huck@wzb.eu

§QuantCo Deutschland GmbH, Am Wasserturm 1, 63110 Rodgau, tobias.schmidt@quantco.com

¶Humboldt University Berlin, Spandauer Str. 1, 10178 Berlin, weizsaecker@hu-berlin.de

JEL classification: D1, D14, D84 , G11

Keywords: Stock market expectations, stock market participation, portfolio choice, financial literacy, complexity

1 Introduction

It is now widely recognized that cognitive abilities can limit the success of financial decision making. One way in which these limitations may arise is through a possible inconsistency between investments and beliefs. Basic models of financial economics prescribe that, by virtue of the investors' rationality, more optimistic beliefs about an asset translate into higher investments in it. Yet the existing evidence for how beliefs about stock market returns drive investments is mainly based on correlations that are, naturally, imperfect measures of the transmission mechanism between beliefs and investments. This paper aims to augment this discussion by considering a sequence of experiments in which participants can earn money in a standard portfolio choice problem that is based on real-world assets. Participants face random variations in returns and our analysis investigates the cognitive restrictions that may impede the effects of these variation.

Specifically, we study a simple portfolio choice problem in a large sample of the German population (a subsample of the SOEP, which is carefully designed for representativeness of the German population) and examine variations of the problem in laboratory experiments. There are three main findings that emerge from the SOEP sample: (i) investment behavior in the experiment has a strong statistical connection to investment behavior in real life, emphasizing the study's external relevance: the average stock market participation rate is 18% in our sample and a one-standard-deviation increase in the experiment's investment predicts an increase in stock market participation by 6 percentage points; (ii) investment choices correlate with stated beliefs as predicted by the standard model; but (iii) neither beliefs nor choices react to exogenous changes in the distribution of returns.

The third finding struck us as deserving of additional enquiry, which is why we moved to variations the laboratory that examine which properties of the choice problem drive the observed deviations from rationality. In particular, we examine whether equivalent variations of the choice problem requiring simpler mental operations would increase best-response rationality. The evidence in favour of this conjecture is our last main finding: (iv) Exogenous changes in the return of a safe asset yield significantly better responses than exogenous changes in the return distributions of a risky asset.[1]

Our findings speak to two large literatures; first, the empirical literature on household finance that finds for many countries, including our test case Germany, puzzlingly low stock market participation rates, and second, the experimental literature on choice under risk and belief biases. In terms of the former literature, our results are largely in line with the previous findings on the correlation of cognitive measures and stock market behavior (for the German context, see Bucher-Koenen & Lusardi, 2011). We add evidence on a new hypothesis: additional incentives to make risky investments—whether they are due to improved market conditions or to policy interventions such as changes in taxation—increase stock market participation only for the privileged subgroups, i.e., the better educated and high earners. As a consequence, caution is due for predictions of the effects of policy interventions that aim to increase stock holdings.

In terms of the literature on risky decision making we believe that we add a genuinely novel result that cannot be explained by standard theories of decision-making under uncertainty but appears to be connected to the psychology of arithmetics. Changes in the incentive structure through a shift of returns induce more rational responses when applied to a safe asset rather than a risky asset. The result suggests the more general effect that people's success in adding a constant to something depends on what this something is. Performing the addition may more generally be relatively easy for a single number and harder for a non-degenerate distribution.[2]

The remainder of the paper is organized as follows. In Section 2 we discuss the related literature in more detail. In Section 3 we describe the experimental

[1]The laboratory data also provide a noteworthy difference to the SOEP sample in that university students's investments *do* react to the variation in incentives. However, they do not show fully rational data patterns not least in that their beliefs, too, react far too little to the variation.

[2]We note here that all incentive shifts in our experiments are presented in the same format. A controlled variation of the shift sizes and a simultaneous variation of an illiquid asset generates isomorphy within pairs of incentive shifts.

design and procedures for both the household panel and the laboratory. In Section 4 we focus on the experimental data and study the relation between beliefs about returns and investments in the experiment. In Section 5 we turn to the validity questions that relate the experimental data to socioeconomic data from the household panel, and in Section 6 we examine the treatment effects. Section 7 presents the additional experiment comparing the return manipulation between safe and risky assets, and Section 8 concludes.

2 Relation to existing literature

The observation that stock market participation is puzzlingly low is widely credited to Haliassos and Bertaut (1995) who find that not only do relatively few members of the middle class invest in stocks, but even amongst the rich, where classical rationales for non-participation are unlikely to hold, participation is far from universal. Germany is a strong case for this puzzle, with its low percentage of stockholders. Behavioral explanations of the puzzle are common in the literature[3] and observational or experimental findings on financial literacy and subjective expectations abound (see e.g. Bucher-Koenen & Lusardi, 2011).

A growing literature measures the general public's beliefs about stock returns. The earlier surveys asked for measure of central tendency only (Vissing-Jorgensen, 2004) whereas entire distributions have subsequently been elicited[4] The survey questions typically ask for statements about the probabilities of market returns lying above given thresholds.[5] The broad picture emerging

[3]Frequently mentioned explanations are education, cognitive skills (Grinblatt, Keloharju, & Linnainmaa, 2011) and financial literacy (van Rooij, Lusardi, & Alessie, 2007), transaction cost and availability of information, and ambiguity aversion (Dimmock, Kouwenberg, Mitchell, & Peijnenburg, 2013).

[4]See the Survey of Economic Expectations (Dominitz & Manski, 2011), the Michigan Survey of Consumers (Dominitz & Manski, 2011), the American Life Panel (Hurd & Rohwedder, 2012), the French 'Mode de vie des Français' panel (Arrondel, Calvo-Pardo, & Tas, 2012) and the Dutch CentER panel (Hurd, van Rooij, & Winter, 2011).

[5]E.g., in the Health and Retirement Survey respondents are asked for the chance that mutual fund shares "will be worth more than they are today" and the chance that "they will have grown by 10 percent or more" (Dominitz & Manski, 2007). Assuming no measurement

from this literature is that expectations are extremely heterogeneous, often lie far away from actual returns (Hurd et al., 2011)[6] and show positive predictive power for stock market investments.

In contrast to previous findings, the respondents in our sample report beliefs that accurately capture the historical market return distribution, at least in the aggregate (see Appendix A). A further notable difference is that while experimental investments have high external validity in our sample, the elicited beliefs have much less predictive power for stock market participation. This may in part be due to the different parts of the sample which enter into the econometric analysis. In Section 5, we report evidence that is consistent with such sample selection. For respondents with a university degree, there *does* exist a positive correlation between stock market beliefs and stock market participation.[7]

While there is a large literature on how people make risky choices[8] and on the relevant correlates,[9] there are no existing studies that we know of that examine whether risky choices in simple lab-style portfolio problems help to predict stock holdings. But while our finding of a strong correlation between an experimental investment and real-life stock market participation is new, the idea is not. In the working paper version of Dohmen et al. (2011) the

error these two questions yield two points on the CDF and, with distributional assumptions, allow fitting an entire distribution.

[6]For example, Kézdi and Willis (2009) find that in 2002 the average subjective probability of a stock market gain was just 49% compared to a historical frequency of 73%. Dominitz and Manski (2011) report that from 2002 to 2004, the average subjective probability of a gain was 46.4%.

[7]But there is further evidence suggestive of a systematic difference between the German sample and others: the subjective probability of the relevant stock market index making a gain varies significantly less between stockholders and non-stockholders in our data than it does in the other studies. In each of Hurd et al. (2011), Dominitz and Manski (2011) and Arrondel et al. (2012), the stockholders assign about ten percentage points more probability mass to the event that the relevant index makes a gain. In our data, this probability differs between stockholder and non-stockholders only by 2.3 percentage points.

[8]For evidence on choice patterns in representative samples, see, e.g. Andersen, Harrison, Lau, and Rutström (2008), Rabin and Weizsäcker (2009), von Gaudecker, van Soest, and Wengström (2011), Huck and Müller (2012) or Choi, Kariv, Müller, and Silverman (2014).

[9]For example, Guiso, Sapienza, and Zingales (2008) show with Dutch household panel data how general trust correlates with stock holdings.

authors report on an investment experiment that was also done in a German household survey but is simpler than ours. Dohmen et al. make the important observation that domain-specific risk attitudes are better predictors of real-world behavior than general risk attitudes. This is consistent with our finding that a choice framed in the context of financial markets is a better predictor for real-life stock holdings than, for example, the respondents' general risk tolerance.

There is also a growing literature on how the complexity of the choice environment can produce suboptimal choices and muted reactions to changes in incentives. Wilcox (1993) and Huck and Weizsäcker (1999) present laboratory experiments showing that complexity of simple lotteries affects lottery choices. Chetty, Looney, and Kroft (2009) show that consumers react to the inclusion of sales taxes on price tags even if the after-tax price of goods does not change. Abeler and Jäger (2015) find much the same thing in a laboratory real-effort task in which earnings are taxed either according to a straightforward schedule or a more complex schedule. Though both schedules yield the same optimal work effort in theory, subjects who face the complex schedule are further away from the optimal solution. Moreover, and similar to our findings, participants with comparatively low cognitive abilities react less strongly to the imposition of new tax rules under the complex schedule.[10]

3 Experimental Design and Procedures

3.1 Survey module

Our experimental module was part of the 2012 wave of the German Socioeconomic Panel's Innovation Sample (SOEP-IS). The SOEP is a nationally representative sample of the German population and the SOEP-IS is its sister survey which is used to try new questions and modules (see Richter &

[10]We note that given the lack of response to stark variations in incentives that we observe in our study, it is perhaps not surprising that investors *do* react to other, extraneous information such as advertisements for individual stocks or photos of financial advisors (Bertrand, Karlan, Mullainathan, Shafir, & Zinman, 2010).

Schupp, 2012, for details). Its sampling of households follows the same procedure as the SOEP does and renders the SOEP-IS approximately representative of the German population. The module was presented to 1146 respondents in 700 households, all of which were added to the SOEP-IS sample in 2012. All households completed the SOEP baseline questionnaire on the same day as our experimental module. Trained interviewers collected responses via computer-aided personal interviewing (CAPI) at the respondents' homes. In the data analysis, we will only use the responses from the "head of household", whom we take to be the household member who responds to the household questionnaire in addition to the personal questionnaire that every household member answers.

Our module contains a regular survey component that we use to elicit several aspects of respondents' asset portfolio (liquid assets, debt, retirement savings) as well as financial literacy and attitudes towards savings and risk.[11] The core component of the module is the interactive experiment modeled on the standard portfolio choice problem that we describe in the following.[12][13]

The first screen of our experiment shows respondents a summary description of the investment decision. They are asked to imagine owning €50,000 that they will invest for the duration of one year. The two available assets are a safe asset that pays 4% and is framed as a German government bond, and a risky asset, referred to as the "fund". The fund is based on the DAX, Germany's prime blue chip stock market index. Respondents receive a one-sentence description of the DAX and learn that, depending on the treatment, the fund pays a return equal to a DAX return drawn from the historical distribution plus a percentage point shifter. There are five treatments that differ

[11] We use Dohmen et al. (2011)'s question, "How willing are you to take risks, in general?", on a scale from 0 to 10. In parts of our analysis we bin the responses into "Risk Tolerance: Low" (response between 0 and 3), "Risk Tolerance: Medium" (4–7) and "Risk Tolerance: High" (8 or above).

[12] In order to minimize interviewer influence, the CAPI-notebooks are placed in front of the respondents and they themselves get to enter their responses. Interviewers are instructed to intervene only if respondents show visible difficulties with the task or explicitly ask for help.

[13] A complete set of instructions is available in the Supplementary Material.

in the value of the shifter, with possible values in the set $\{-10, -5, 0, 5, 10\}$. Respondents are randomly allocated to treatments. If their shifter value is 0, then the shifter is not mentioned (for simplicity). Otherwise the first screen indicates the absolute size of the shifter but not its sign. For example, a respondent would learn that the fund pays either 5 percentage points less than the DAX or 5 percentage points more than the DAX and that she will subsequently learn which of the two values applies. The respondents also learn that they will be paid in cash on a smaller scale at the end of the survey.

On the second screen, respondents receive more detailed explanations about the determination of payments including (in bold letters) the information of the shifter's sign that "the computer has determined through a random draw". We use this two-step revelation of the shifter's random draw in order to maximize the respondent's appreciation that the shifter is random with zero mean, carrying no information about the underlying DAX return. Since each respondent is only confronted with one realized shifter value in their choice problem, showing the mirrored value makes it salient that the shifter carries no information. The procedure also ensures that the instructions of the laboratory replication are identical despite the fact that only two shifter values are possible there (see Section 3.2 below).

The text on the second screen also gives some numerical examples and specifies that the fund's return depends on a draw from historical DAX returns from 1951 to 2010 and that actual payments are scaled down by a factor of 2000.[14]

Upon reading these short instructions the respondents make their investment decision on the third screen. Respondents who invest their entire endowment in the riskless asset would receive a certain payment of €26. Investing the entirety in the risky asset could yield a payment anywhere from €11.52 to

[14]For all years since the DAX's origination in 1988 we use the actual yearly returns on the index. For all previous years we make use of the yearly return series from Stehle, Huber, and Maier (1996) and Stehle, Wulff, and Richter (1999), who impute the index going back all the way to 1948. All returns are nominal. In contrast to e.g. the S&P 500 the DAX is a performance index, which means that dividend payments are included in the return calculations.

€56.52 depending on the treatment and the randomly drawn year. No information on historical returns is made available to the respondents during the experiment. Under the assumptions of rational expectations, EU-CRRA and usual degrees of risk aversion, one can generate the approximate prediction that in treatments with non-negative shifters, all respondents with degree of relative risk aversion below 3 should invest their entire endowment in the risky asset; those with a shifter of -10 should invest very little whereas those with -5 should invest intermediate amounts.[15]

On the fourth screen we elicit respondents' beliefs about the return of the fund, using the histogram elicitation method pioneered by Delavande and Rohwedder (2008) and refined by Delavande, Giné, and McKenzie (2011) and Rothschild (2012).[16] A screenshot of the interface can be found in Online Appendix F. Respondents have to place 20 "bricks", each representing a probability mass of 5%, into seven bins of possible percentage returns.

[15]These statements hold in a classic two-period two-asset portfolio choice model with log-normal asset returns and CRRA utility over wealth in the second period (i.e. a simplified version of Merton (1969) and Samuelson (1969); see also Campbell and Viceira (2002)). In this model the optimal stock investment share α can be approximated by

$$\alpha = \frac{\mu_r - r_f + \sigma_r^2/2}{\rho \cdot \sigma_r^2},$$

where μ_r is the expected log return, σ_r^2 is the variance of returns, r_f is the natural logarithm of the risk-free rate and ρ is the coefficient of relative risk aversion. Over the payoff-relevant period 1951-2010 the log-normality assumption was approximately correct for year-on-year returns on the DAX (Shapiro test p-value: 0.6), the mean log-return was 0.11 and the variance of returns was 0.1. The riskless asset in the experiment paid 4%. The predictions made in the main text readily result under rational expectations. For respondents with log-utility ($\rho \approx 1$) the optimal stock investment share in Treatment 0 is 1, in Treatment -5 it is 0.74 and in Treatment -10 it is 0.22. Under the same assumptions positive shifters have no effect on stock investment, which remains at the corner solution. However, given that stock investments observed in reality are often much lower than those predicted by the model and that most of the finance literature estimates risk aversion to be substantially higher we decided to also include positive shifters.

[16]For an overview of studies which have used this or similar methods see Goldstein and Rothschild (2014) and references therein. A popular alternative method for the elicitation of a distribution is to ask for subjective probabilities of surpassing given thresholds. One drawback of that methods is that responses are often internally inconsistent (Binswanger & Salm, 2013). In the Health and Retirement Survey 41% of respondents give the same answer to both the question about the likelihood of a positive return and the question about a return above 10%, and a further 15 % violate monotonicity.

The set of available bins is {(-90%,-60%),(-60%,-30%),(-30%,0%),(0%,30%),(30%,60%),(60%,90%),(90%,120%)}. The bins are, hence, wide enough to allow responses over the entire historical support of DAX returns[17] and, more generally, allow for a large set of possible subjective beliefs. In addition, on the fifth screen, respondents enter the "average return [they] expect for the fund". For both the histogram elicitation of beliefs and for the stated beliefs, it is straightforward to formulate the rational prediction of treatment differences: no matter the distribution of beliefs in the population, the shifter should move beliefs one-to-one. For example, between the -10 shifter and the +10 shifter treatments reported beliefs should differ by 20 percentage points.

Like all previous surveys on beliefs about stock market returns we decided not to incentivise either of these belief measures. Properly incentivising the answers would have required a payment mechanism whose explanation would have strained the attention span of our respondents (see Allen, 1987, for an example of such a mechanism) and taken up valuable survey time for very little gain.[18]

On the sixth and seventh screens, respondents report how confident they are of their belief statements, on a scale from 0 ("not at all") to 10 ("very sure"), and answer a few understanding questions. The eighth screen elicits the respondents' beliefs about next year's DAX return using the same histogram interface that was used before. Finally, on the ninth and last screen of the experimental module respondents were told which of the years between 1951 and 2010 had been drawn and received a detailed calculation for their payment. Respondents were paid in cash, with amounts rounded up to the nearest euro, at the end of the entire survey interview. On average respondents received

[17]The lowest return on the DAX in the payoff-relevant period was -43.9% in 2002. The highest return was 116.1% in 1951. The lowest bin was included for reasons of rough symmetry and to keep subjects from anchoring their reports on the lowest possible return displayed in the interface.

[18]Both Armantier and Treich (2013) and Trautmann and van de Kuilen (2015) show that the wrong scoring rule can induce bias in the responses. In contrast, not incentivizing the elicitation of beliefs does not yield biased answers in these studies but merely noisier answers. A further concern with incentives is the introduction of possible motives for attempted hedging between tasks (see e.g. Karni & Safra, 1995).

	Dependent variable: Participation in the Experiment
Female	-0.001 (0.031)
Born in the GDR	0.028 (0.041)
Abitur	0.043 (0.056)
University Degree	-0.001 (0.069)
Household Size	-0.018 (0.020)
Number of Children in Household	0.019 (0.035)
Employed	0.017 (0.037)
Financially Literate	0.028 (0.031)
Interest: < 250 Euros	-0.028 (0.034)
Interest: 250 - 1.000 Euros	0.027 (0.051)
Interest: 1.000 - 2.500 Euros	0.096 (0.100)
Interest: > 2.500 Euros	0.120 (0.240)
Interest: refused to answer	-0.076 (0.086)
Stock Market Participant	0.025 (0.047)
Risk Tolerance: Low	0.029 (0.033)
Risk Tolerance: High	0.027 (0.044)
Age bracket 31-40	0.032 (0.086)
Age bracket 41-50	-0.083 (0.069)
Age bracket 51-60	-0.084 (0.068)
Age bracket 61-70	-0.064 (0.067)
Age bracket > 70	-0.200^{***} (0.068)
N	692

*p < .1; **p < .05; ***p < .01
Standard errors are bootstrapped with 1000 replicates

"Financially Literate" is an indicator variable which is 1 whenever the respondent states that he/she is either "good" or "very good" with financial matters. For details on this and the other variables, see Online Appendix H.

Table 1: Selection into the experiment: Probit marginal effects

€27.16 (min: €17, s.d.: €3.43, max: €48).[19]

Before respondents are presented with the experimental module and its instructions, they have a choice whether or not to participate. The participation rate is 80%. Those who decline primarily cite old age and problems with using computers but also a lack of interest in financial matters or ethical or religious reservations against any sort of financial "gambling". The probit regression shown in Table 1 mirrors these answers from the open-ended question about the reasons for non-participation. The most potent predictor, indeed the only

[19] Had they invested everything into the fund the average earning would have been €28.88 resembling a return of 15.5% – more than three times as much as the safe return of 4%.

significant predictor, of selection into the experiment is age. Respondents over the age of 40 are somewhat less likely to participate and respondents above the age of 70 are significantly less likely to participate though almost two thirds in this age group still participate. All other observable characteristics play no role in the selection into the experiment. A Wald-test for the joint significance of all variables other than the age brackets cannot reject the null of no effect ($\chi^2(18) = 19.41$, $p = 0.37$).

3.2 Laboratory Experiment

Upon completion of the field data collection in the SOEP-IS, we used the identical experimental module for a set of 198 university students in the WZB-TU Berlin decision laboratory. Recruitment into the laboratory sample followed standard procedures.[20] The instructions and sequence of informational displays on the computer screens in the laboratory were as close to the CAPI environment as we could produce them, so that the potential practical difficulties with the format would affect both populations. The experimental participants' payments were also scaled by the same factor as payments to SOEP participants. The only relevant difference in experimental design and procedures are that (i) the experimental participants do not have to fill out the long SOEP questionnaire, and (ii) we conducted only two treatments with return shifters -10 and 10, in the laboratory, focusing on the strongest treatment difference in incentives. Since the SOEP respondents who happened to be in either of these two treatments were only informed about the existence of these two treatments, we could leave the instructions entirely unchanged between survey and lab environments.

[20]The decision laboratory uses ORSEE (Greiner, 2015).

4 Experimental Data

4.1 Beliefs and Investments

We start with a summary description of investments and elicited beliefs. We call the share of wealth that a respondent invests in the fund "equity share" hereafter. In both samples the distributions of equity shares have relatively wide supports and few people invest all or nothing. Summing over all treatments, the means (and standard deviations in parentheses) of the equity share are 0.37 (0.25) in the SOEP sample and 0.46 (0.31) in the laboratory sample. The proportions of respondents investing all, exactly half, or nothing in the risky asset are 0.03, 0.2 and 0.18 in the SOEP sample and 0.12, 0.05 and 0.09 in the laboratory sample.[21] 82% of the SOEP respondents invest in the risky asset, which is much higher than the actual stock market participation of around 18%. The stark difference may be due to the large salience and availability of the risky asset and/or it may indicate an experimenter demand effect. The greater extensive margin of investment works in favour of performing statistical analyses.

A description of the beliefs about the fund's return is more involved, since each belief report consists of an entire histogram. A clear difference between the SOEP and the lab is that the laboratory participants use more bins than the SOEP respondents.[22] The median number of bins that contain at least one brick is 6 in the laboratory while it is only 3 in the SOEP where 28% of respondents use only a single bin and a further 14% only use two bins.[23]

In the analysis below we repeatedly use summary statistics that we compute from the reported histograms. Using the stated point beliefs would produce similar results in most instances, and we will often present the results of both measures, pointing out the differences where they arise. To compute

[21] When we restrict the SOEP sample to the two extreme treatments that we also ran in the lab the proportions are 0.05, 0.21 and 0.15.

[22] Appendix G contains examples of the raw data of elicited histograms from both samples.

[23] Relative to comparable studies that use similar methods, the mentioned frequencies are on the low side. Delavande and Rohwedder (2008) report that 73% of their subjects used two or fewer bins.

	Equity Share		Imputed Expectation of Belief		Imputed S.D. of Belief		Stated Expectation of Belief		
	Mean	S.D	Mean	S.D	Mean	S.D	Mean	S.D	N
Overall	0.37	(0.25)	12.53	(20.59)	23.96	(16.54)	8.27	(17.84)	562
Age Bracket									
<30	0.41	(0.27)	12.16	(16.06)	30.25	(16.07)	8.74	(16.64)	82
31-40	0.39	(0.22)	13.85	(15.73)	25.60	(17.13)	12.02	(16.54)	76
41-50	0.40	(0.23)	12.57	(24.70)	26.36	(16.75)	7.12	(18.65)	107
51-60	0.37	(0.26)	13.24	(21.86)	22.72	(16.46)	8.43	(19.41)	107
61-70	0.34	(0.26)	10.02	(19.63)	20.46	(15.88)	6.22	(17.27)	111
>70	0.32	(0.28)	14.13	(22.49)	19.19	(14.77)	8.36	(17.63)	79
Gender									
female	0.35	(0.24)	9.72	(22.29)	25.60	(17.20)	7.86	(21.59)	271
male	0.39	(0.26)	15.14	(18.52)	22.43	(15.78)	8.65	(13.46)	291
Born in									
West Germany	0.37	(0.26)	12.11	(20.97)	23.34	(15.60)	7.40	(17.38)	379
East Germany	0.34	(0.23)	12.87	(21.96)	22.47	(17.46)	7.75	(17.69)	116
abroad	0.42	(0.28)	14.95	(15.44)	29.74	(19.10)	14.66	(17.35)	54
Abitur									
yes	0.37	(0.28)	10.74	(19.51)	26.70	(14.83)	6.40	(13.47)	122
no	0.37	(0.25)	13.02	(20.87)	23.20	(16.93)	8.78	(18.85)	440
University Education									
yes	0.35	(0.28)	11.54	(21.78)	26.95	(15.40)	5.55	(16.46)	72
no	0.37	(0.25)	12.67	(20.42)	23.52	(16.67)	8.67	(18.01)	490
Employed									
yes	0.39	(0.25)	13.64	(20.70)	24.38	(16.13)	8.98	(16.13)	297
no	0.35	(0.26)	11.27	(20.42)	23.49	(17.01)	7.47	(19.58)	265
Financially Literate									
yes	0.36	(0.25)	14.13	(20.80)	24.02	(15.98)	8.08	(17.68)	283
no	0.38	(0.26)	11.05	(20.27)	24.00	(17.14)	8.47	(18.09)	277
Stock Owner									
yes	0.45	(0.29)	12.79	(18.20)	22.66	(14.55)	8.95	(13.82)	107
no	0.35	(0.24)	12.50	(21.13)	24.29	(16.99)	8.11	(18.69)	454

"Financially Literate" is an indicator variable which is 1 whenever the respondent states that he/she is either "good" or "very good" with financial matters. For details on this and the other variables, see Online Appendix H.

Table 2: Experimental Responses in the SOEP by subgroup

statistics like the expectation or the standard deviation of returns from the histograms, we take the 8 points on the CDF, interpolate between them using a cubic spline and then calculate the statistics numerically.[24] Using these imputed distributions, we find that the average of the SOEP respondents' mean expected return of the fund is 12.5% and the average standard deviation of the fund's return distribution is 20.6%. For the laboratory sample, the average mean belief about the fund's return is 11.6% and the average standard deviation is 35.6%.

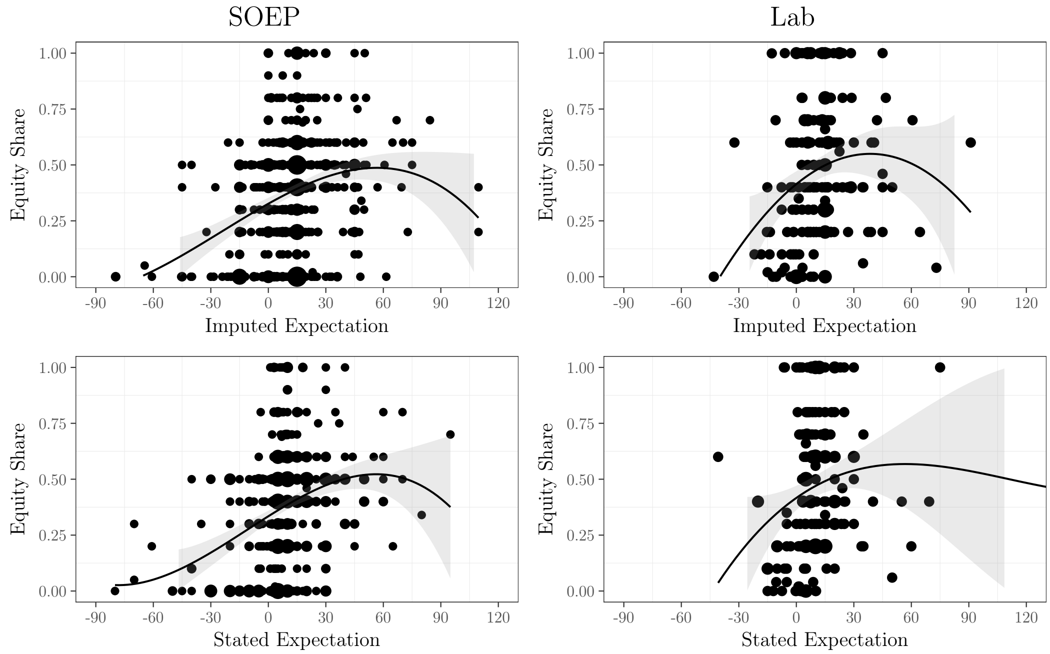
As described in the previous section, we also elicited scalar belief reports by asking for the "expected" fund return. In the SOEP sample, this variable has a mean of 8.3% and a standard deviation of 17.8%. In the laboratory sample, the mean is 11.0% and the standard deviation is 19.1%. Stated expectations are highly correlated with expectations inferred from belief distributions (Spearman correlation coefficient: 0.43 for the SOEP and 0.47 for the lab sample). Table 2 collects key descriptives for the main experimental variables for different subgroups of the SOEP sample (a similar table for the lab sample is omitted because the student population is very demographically homogeneous). The table shows some but not many differences between subgroups. Respondents who are succesful in our financial literacy tests invest, surprisingly, the same average amount in the artificial stock market. Real-world stock owners, however, invest significantly more in it and show only slightly more optimistic beliefs (see also Section 5).

We now investigate the extent to which equity share and beliefs are correlated. Figure 1 contains a scatter plot of equity shares and the belief measures for both the SOEP and the lab sample. The figure shows pronounced positive relationships between belief and investment overall. At the mean of the data an increase in the expected return by one percentage point is associated with a one third percentage point increase in the equity share (see Table A1 in the Appendix for OLS regressions). This relationship holds for both our belief measures and is roughly the same in the laboratory. This evidence of a

[24] This method is due to Bellemare, Bissonnette, and Kröger (2012). A more detailed description of the interpolation procedure can be found in Appendix F.

positive association between beliefs and investments is consistent with many studies in the belief elicitation literature (see, for example, Naef and Schupp (2009) and Costa-Gomes, Huck, and Weizsäcker (2014) in the context of trust games).

As decribed earlier, investments are incentivised but beliefs are not. This may affect the precision of stating beliefs and thus our measurement of correlates of beliefs and investment. To account for such possible effects, Appendix D provides a detailed subsample analysis for subgroups of different degrees of measurement error. The analysis shows that the results presented in the main text are robust to these subgroup restrictions, unless otherwise noted in the main text. We again note that simple models of portfolio choice, see for instance footnote 15, would predict a stronger relationship between beliefs and investments, which is not surprising given that they leave out important factors of the decision process. But given the robustness of our subgroup analysis, measurement error appears not to be the sole reason for the deviation from theory.



Overlapping observations are aggregated, with the dot's size being proportional to the number of observations thus aggregated. Model fit comes from a polynomial regression in which investments are a cubic function of expected return (Models 2, 5, 8 and 10 in Table A1 in the Appendix, which also contains alternative specification that e.g. control for personal characteristics but all show results that are qualitatively and quantitatively similar.). 95% confidence interval in light gray.

Figure 1: Equity Share and Beliefs

Notice that also in other ways, the data show patterns that are hard to square with the predictions of the standard model. As in Merkle and Weber (2014), there is a substantial fraction of participants who expect a negative excess return for the experimental asset and yet invest positive amounts. But altogether, the strong statistical connection between belief data and investment decisions can be regarded as supporting the basic implication of the standard portfolio choice model: higher expected returns occur together with larger investments.

5 External validity: Stock market participation

We now ask which of our response variables are correlated with real-life investments. Specifically, we test the external validity of our data by comparing

Stock-market participation rate by...	Decile									
	1st	2nd	3th	4th	5th	6th	7th	8th	9th	10th
Household Income	7%	7%	3%	21%	14%	17%	20%	19%	26%	46%
Liquid Wealth	0%	2%	2%	2%	5%	13%	11%	39%	43%	56%

Table 3: Stock-market participation rate by income and wealth deciles

elicited behavior in the experiment with survey responses to the question "Do you own any stock market mutual funds, stocks or reverse convertible bonds?"

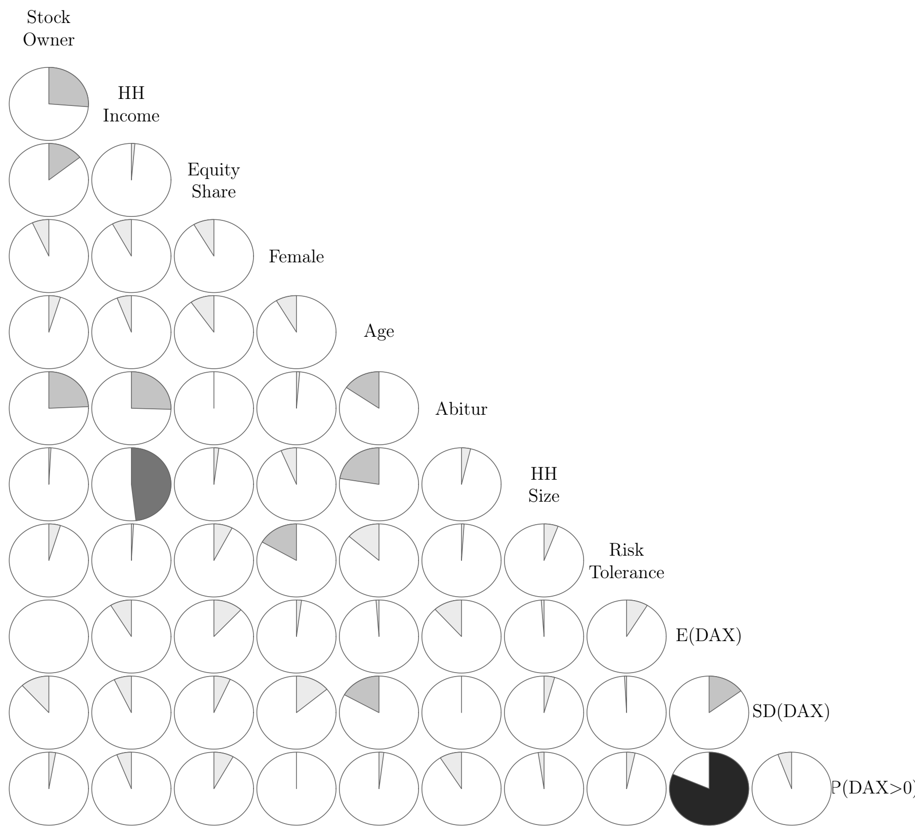
18% of all households answered this question in the affirmative, which is in line with other evidence on the German stock market participation.[25] Splitting the participation rate by deciles of both household income and a proxy for liquid wealth,[26] Table 3 also shows that stock market participation increases in both variables but stays well below 100%.

Figure 2 displays a correlogram, a visualization of the correlation matrix for several survey and experimental variables. Starting from the vertical, positive correlations are displayed as wedges that are shaded clockwise while negative correlations are shaded counter-clockwise. The higher the correlation, the larger the wedge and the darker the shade of the wedge.

The correlogram shows that only a handful of variables are reliable correlates of stock market participation. Most of the significant correlations have been observed in the previous literature. For example, household size is known to be a significant predictor of stock market holdings. Likewise, household income and Abitur—the highest form of secondary education in Germany and the only form that grants access to the university system—are well-known and entirely unsurprising correlates of stock ownership. Notably in our data, equity share is the only *experimental* variable that is significantly correlated

[25] Most other surveys provide numbers only for the percentage of individuals who hold stocks. In our data this percentage stands at 15.4% (S.E.: 1.1%) while a 2012 survey by Deutsches Aktieninstitut (2012) puts it at 13.7%.

[26] The SOEP question about interest earned on investments over the previous year is answered by far more people than more detailed questions about the amounts of wealth held in the form of various assets. We therefore use this variable as a proxy for liquid wealth. The alternative measure, the sum over all asset classes, yields broadly similar results. For details on these variables, see Online Appendix H.



The correlogram above visualizes the pairwise (Pearson) correlation coefficients of the variables.
E(DAX) is the imputed expected return on the DAX going forward while SD(DAX) is the imputed standard deviation of the reported return distribution. P(DAX>0) is the reported probability that the DAX will make a gain over the next year.

Figure 2: Correlogram

with stock holdings (correlation: 0.14, p-value: < 0.001), an observation that is consistent with the hypothesis that the standard portfolio choice problem captures essential aspects of stock market participation, but which could also stem from spill-overs from real-life decisions into the experiment. The elicited beliefs, in contrast, show only weak correlations with stock ownership. Only when interacting with university-degree status (as a proxy of numerical literacy) do we find a significant coefficient of belief, for those respondents with a university degree.[27] Of course, from a theory stand point there may be no

[27]The size of the results depends on whether we use the stated point belief or the average

strong reason for beliefs about the past to impact on current stock holdings.[28]

The correlograms only show bivariate relations. In order to gain a broader picture we investigate whether the correlations change if we control for other variables. We find that equity share has explanatory power over and above the other variables, see Table 4. Even after including all relevant controls, which drives up the R^2 to around 30%, the coefficient for equity share remains both economically important and statistically significant and is robust to different specifications. Back-of-the-envelope calculations yield the result mentioned in the introduction, that an increase in equity share by one standard deviation is associated with an increase in stock market participation of six percentage points.[29]

The fact that equity share helps to explain stock holdings even if we control for all other variables that are known to be good predictors of stock market participation is important for two reasons. First, it establishes external validity. Investment behavior in the experiment is strongly related to investment behavior outside of the experiment. Second, the result gives hope that the simple experimental portfolio choice problem can be used as a simple piloting device: it allows the controlled manipulation of a behavioral variable that has a close connection to stock market particpation, both in terms of economic theory and in terms of empirical correlation. Hence, there is hope that interventions, for example, to encourage stock ownership, could be pre-tested in laboratory or artefactual field experiments such as ours.

"imputed" belief from the respondent's histograms. A unit standard deviation increase in the stated expectation about the fund is associated with a stock ownership increase of 11 percentage points (p=0.001, one-sided t-test) while a unit standard deviation increase in the imputed expectation about the fund is associated with a stock ownership increase of 6 percentage points (p=0.11).

[28]The results of this section are robust against replacing backward looking beliefs by forward beliefs about the development of the DAX in the next twelve months. A notable (reasonable) difference is that forward looking beliefs are less strongly correlated with equity share, for which backward looking beliefs are more relevant.

[29]Appendix D shows that also for respondents who show low measurement error equity share predicts stock ownership better than expectations predicts stock ownership.

	Dependent variable: Stock Market Participant		
	(1)	(2)	(3)
Equity Share	0.220***	0.240***	0.200***
	(0.072)	(0.068)	(0.064)
Female		−0.043	−0.029
		(0.032)	(0.030)
Born in East Germany		−0.058*	−0.044
		(0.034)	(0.033)
Age		0.006	0.004
		(0.005)	(0.006)
Age^2		−0.0001	−0.0001
		(0.0001)	(0.0001)
Abitur		0.200***	0.150**
		(0.061)	(0.058)
University Degree		0.049	−0.003
		(0.078)	(0.072)
Household Size		0.039**	−0.004
		(0.019)	(0.022)
Risk Tolerance: Low		0.020	0.034
		(0.037)	(0.035)
Risk Tolerance: High		0.008	0.058
		(0.044)	(0.043)
Imputed expectation of DAX		0.001	0.0003
		(0.001)	(0.001)
Imputed S.D. of DAX		−0.003***	−0.001
		(0.001)	(0.001)
Gain Probability of DAX		−0.003	0.039
		(0.088)	(0.085)
Number of Children in Household		−0.096***	−0.057*
		(0.030)	(0.030)
Employed		−0.015	−0.024
		(0.036)	(0.037)
Financially Literate		0.140***	0.080***
		(0.032)	(0.031)
Interest: < 250 Euros			0.061*
			(0.033)
Interest: 250 - 1.000 Euros			0.270***
			(0.057)
Interest: 1.000 - 2.500 Euros			0.430***
			(0.086)
Interest: > 2.500 Euros			0.310***
			(0.110)
Interest: refused to answer			0.150
			(0.100)
Household Income (missing=0)			0.023
			(0.018)
Household Income: missing			0.210**
			(0.084)
Constant	0.110***	−0.130	−0.130
	(0.029)	(0.140)	(0.140)
N	561	560	560
R^2	0.021	0.150	0.280
Adjusted R^2	0.019	0.130	0.250

*p < .1; **p < .05; ***p < .01
Household income is in thousands of Euros

Household income is set to zero where missing (48 cases). Moreover, a dummy variable is added to the regression which is 1 for the observations with missing household income.

Table 4: Predicting real-world stock-market participation

6 Treatment effects

Recall that we implement five exogenous treatments that shift the historical return of the DAX. The shifts are sizable, ranging from -10 percentage points to +10 percentage points. Table 5 documents that, by and large, there is no sizable effect of the return shifter on equity share in the SOEP sample (see also Online Appendix G showing histograms of equity shares in the different treatments). The lack of response can hardly be explained by small incentives. In terms of the nominal framing of the €50,000 investment, the difference in returns between Treatments -10 and 10 amounts to a difference in returns of up to €10,000. In terms of the real monetary value of the experimental investment, the variation in return amounts to a difference of up to €5. This difference is large enough for the typical participant in an experiment (even in representative samples) to react. The overall lack of response therefore suggests that many respondents find it difficult to incorporate the shift appropriately in their investment choice.

However, this result is not universal. Instead Table 5 shows an important difference between the SOEP and the laboratory sample. While SOEP participants appear to ignore the shifter on average, there is a strong and statistically significant reaction of investments to the treatment in the laboratory. There, the equity share rises from 0.30 to 0.63 in response to improving the return of the fund by 20 percentage points.

Similar results hold for those parts of the SOEP sample that are plausibly more financially savvy, those who are more educated, those who have more liquid assets (or refuse to answer the question about how much interest they obtain from liquid assets) and those who answer the standard financial literacy question about compound interest correctly. Hence, it appears that the main difference between SOEP and lab is driven by selection on educational covariates and wealth.[30]

The beliefs about the fund's return, however, do not respond to the shifter in the way they should, no matter what measure of beliefs we use and no matter

[30]For details of differences between subsamples, see Appendix C.

Setting	Variable	-10	-5	0	5	10	ANOVA	Kruskall-Wallis
SOEP	Equity Share	0.40 (0.02)	0.34 (0.02)	0.32 (0.02)	0.39 (0.02)	0.39 (0.02)	0.106	0.135
	Imputed Beliefs	13.14 (1.97)	10.58 (1.81)	9.38 (1.85)	14.48 (1.83)	14.45 (2.18)	0.232	0.326
	Stated Beliefs	8.55 (1.71)	7.68 (1.70)	6.60 (1.98)	9.28 (1.43)	8.93 (1.66)	0.810	0.990
	Probability of a Gain	0.68 (0.03)	0.67 (0.03)	0.67 (0.03)	0.74 (0.02)	0.69 (0.03)	0.323	0.313
Lab	Equity Share	0.30 (0.03)				0.63 (0.03)	0.000	0.000
	Imputed Beliefs	10.05 (1.71)				13.37 (1.57)	0.156	0.016
	Stated Beliefs	9.87 (2.28)				12.30 (1.38)	0.374	0.004
	Probability of a Gain	0.59 (0.02)				0.65 (0.01)	0.029	0.009

Table 5: Mean levels by treatment

whether we consider the SOEP data or the laboratory data. While there is a statistically significant effect in the laboratory sample, it is much smaller than the 20 percentage points predicted by probabilistic sophistication, and there is no effect at all in the SOEP sample. In both samples and regardless of whether we consider imputed beliefs or stated beliefs, we can strongly reject the rational prediction that the shifter moves the mean of beliefs one-to-one.

As we show in Appendix A, the participants' beliefs about past DAX returns are surprisingly accurate. Within each of the seven histogram bins, the population-average belief of DAX returns falling in the bin is within just few percentage points of the historical frequency. But as described in the previous paragraph, the beliefs do not react strongly enough to the experimental manipulation.

We tentatively conclude from the experiments that human decision makers, despite judging a risky return distribution well, may be unable to deal with manipulations of it well. This raises the question how well the respondents understand the manipulation, despite our long and intense efforts for clarity in the instructions. The next section investigates the possibility that the weak reaction to the manipulation may be driven by factors beyond the understanding of the experimental instructions.

7 Asset Complexity and Reactions to Changes in Incentives

7.1 Experimental Design

In this section, we investigate the role of complexity with an additional laboratory experiment. We introduce manipulations of both the risky asset and the safe asset that are economically equivalent and described in identical terms. To make the two shifts economically equivalent, we modify the decision maker's exogenous income level.[31]

The design follows the same format as the paper's main experiment, implementing the standard portfolio choice problem. In the new experiment (i) each participant makes eight investment decisions, allowing a within-subject analysis, and (ii) each participant receives a task-specific fixed income in addition to the earnings from the portfolio choice.

The participants are endowed with an illiquid asset that generates the fixed income W_I, and with liquid wealth W_L that they can allocate among a safe asset and a risky asset. The risky asset pays a rate of return r whereas the safe asset pays a rate of return r_f.

Now consider an increase in the risky return r by an amount Δ, analogous to the exogenous return manipulation of the paper's main experiment. Under this manipulation, a decision maker who invests α in the risky asset earns a random payoff given by:

$$\pi(\alpha) = \alpha W_L(1 + r + \Delta) + (1 - \alpha)W_L(1 + r_f) + W_I$$

For a framing variation of this manipulation by Δ, we can alternatively induce a simultaneous shift in r_f by amount $-\Delta$ and in W_I by amount ΔW_L, yielding the same payoff from investing a share α in the risky asset:

[31] Two remarks are in order. First, we designed this section's experiment after we observed the results from the experiments described in Section 3.2—hence the separate presentation. Second, the fact that we could run the complexity experiment only in a laboratory format also means that we cannot investigate the present research question for the subsamples that show the weakest reaction to incentive shifts. We suspect, but have no proof, that these subsamples would exhibit even larger differences in their reactions to different shifts.

$$\pi(\alpha) = \alpha W_L(1 + r) + (1 - \alpha)W_L(1 + r_f - \Delta) + (W_I + \Delta W_L)$$

From the fact that $\pi(\alpha)$ is identical between both treatments and for all α, we conclude that the same risks are available between the two manipulations. Consequently, expected utility theory, and any other theory that employs a stable mapping from a constant set of uncertainty states into the risky asset's return rate, predict an identical choice by the decision maker. The same statement is true if both the safe and the risky assets' returns are additionally shifted by a constant amount Δ'. The experiment's null hypothesis is thus that participants react equally between the equivalent manipulations of incentives applying to the safe asset or the risky asset.

To ensure that the results are not driven by an asymmetry between positive shifts and negative shifts, we formulate the entire experiment such that only positive shifts occur. This is achieved by adding an appropriate return shift Δ' to both assets.[32] The parameters for the eight choice problems are displayed in Table 6. The collection of equivalent variations is the following: Problems 1 and 3 are economically equivalent, Problems 2 and 4 are economically equivalent, Problems 5 and 7 are economically equivalent, and Problems 6 and 8 are economically equivalent. Problems 1 and 2 differ only in the risky asset's return; Problems 3 and 4 differ in the shifter applied to the riskless asset (and a compensatory change in the illiquid endowment), in the described way. But the difference in incentives is the same between 1 and 2 as between 3 and 4. Thus, expected utility and most of its generalizations predict that the difference in investments is identical. Analogously, the difference between 5 and 6 is predicted to be identical to the difference in investments between 7 and 8. As described above, our main hypothesis in this experiment is that shifts in

[32] We also ran three pilot sessions but do not use the data gathered in these sessions here. In the first pilot session subjects were presented with both "bonuses" and "fees" on the two assets and displayed aversive reactions to any asset to which a fee was applied. Since the effect of gain/loss framing was not the subject of this study we therefore ran two sessions with bonuses only but found that up to 42% of subjects chose investments at the lower boundary of the budget set. Since this much truncation presents problems both in terms of power and in terms of the distributional assumptions one is required to make to deal with it, we therefore changed the magnitude of the bonuses to arrive at the valued reported here, values that yield much fewer truncated responses. Note, however, that the responses in all pilots were also indicative of stronger reactions to changes in the safe asset.

Treatment	Bonus on Risky Asset	Bonus on Riskless Asset	Illiquid Endowment	Liquid Endowment
1	9.00	5.90	16000	50000
2	2.65	5.90	16000	50000
3	5.90	2.80	17550	50000
4	5.90	9.15	14375	50000
5	9.10	6.05	14275	50000
6	3.10	6.05	17275	50000
7	6.05	3.00	15800	50000
8	6.05	9.00	15800	50000

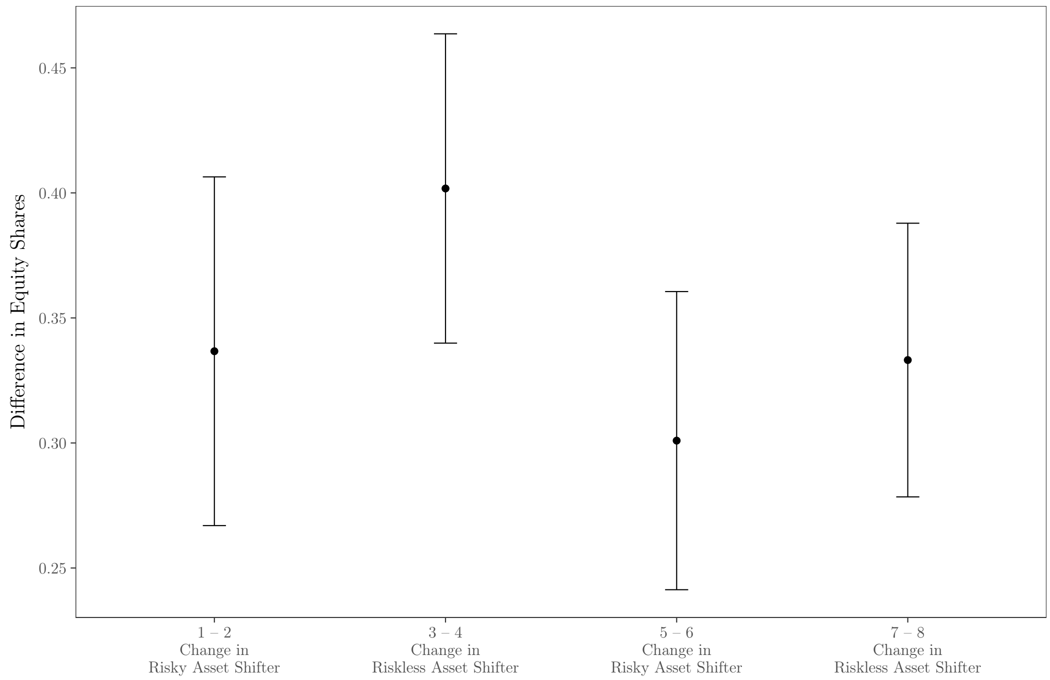
Table 6: Treatment parameters

safe return generate a stronger reaction: investments may differ more between 3 and 4 than between 1 and 2, and more between 7 and 8 than between 5 and 6.

76 participants were recruited into 4 experimental sessions at WZB-TU Berlin laboratory in the spring of 2014, using identical procedures as in the study described in Section 3.2. Similar to the first lab study we take a fixed-interest German government bond (here, yielding 2 % per annum) as the safe asset and the return on the DAX in a year randomly drawn from 1951 to 2010 as the risky asset. Treatments were presented in random order so as to avoid confounds from learning or contrast effects. One of the eight tasks was randomly selected and paid out at the end of the experiment, ensuring incentive compatibility for each task.

7.2 Results

Figure 3 displays the differences in average equity shares (the percentage of the liquid endowment invested in the risky asset) for each of the four treatment pairs. A weaker reaction to changes in the risky asset return is immediately visible. Treatments 1 and 2 vary the risky asset return by 6.35 percentage points while holding the riskless asset return constant. This causes a change in mean equity share from 0.28 when the bonus on the risky asset is 2.65



Error bars show 95% confidence intervals.

Figure 3: Investments in the risky asset by treatment

percentage points to 0.62 when the bonus on the risky asset is 9 percentage points, for a difference of 0.34. A change of equal magnitude in the return of the riskless asset causes a larger change in the equity share. While the mean equity share in treatment 3 is 0.61, almost identical to that in treatment 1, the mean equity share in treatment 4 is 0.21, lower than that in treatment 2. This yields a difference of 0.4. The same pattern of responses hold analogously for treatments 5 to 8.[33]

Given the comparatively small sample size, each of these mean responses is subject to considerable sampling error. In order to formally test our main hypothesis we compute the difference in differences for treatments 1 to 4 and add to this the difference in differences for treatments 5 to 8 (this form of pooling preserves independence among individuals). Under the null of rational,

[33]A graph of the raw responses is available in Online Appendix L.

equal-sized responses to changes in either the risky and riskless asset returns this sum should be zero. Instead, we find it to be 0.10, positive and statistically significantly so (one-sided Wilcoxon rank sum test p-value = 0.014, one-sided t-test p-value = 0.047).[34] In all, it appears that changes in the riskless asset are easier to process (or understand) than shifts in the risky asset.

8 Conclusion

The paper at hand describes a simple portfolio choice problem with one safe and one risky asset, implemented in an artefactual field experiment for a large population sample in Germany. The data from this experiment exhibit high degrees of external validity between certain variables inside and outside the experiment. In this sense the choice problem, despite its extreme reduction, captures important real-life tradeoffs in financial markets. We also find that households are remarkably unresponsive to shifts in returns.

The more detailed analysis also shows that the degree of external validity, i.e., the extent to which our results help to predict actual stock market participation, varies between different subgroups. External validity is stronger for skilled and savvy subjects. We also observe that only these savvier subgroups of subjects respond in a meaningful way to changes in incentives, highlighting, once again, the important role of cognitive ability for even the simplest of financial decision problems (Benjamin, Brown, & Shapiro, 2013). In our setting less educated subjects forgo substantial additional earnings by not responding to exogenous shifts in investment incentives. Related to previous studies on financial literacy (e.g. Lusardi and Mitchell (2011) on retirement savings, Gerardi, Goette, and Meier (2013) on mortgage foreclosure and von Gaudecker (2015) on portfolio diversification), this difference addresses the possibility of distributional effects that arise from cognitive differences. Similar interventions to foster investments in real life (such as tax subsidies for

[34] Over all treatments about 11% of responses are truncated below at zero. The percentage of truncated responses is higher in treatments 4 and 8 than it is in treatments 2 and 6. The truncation therefore potentially obscures larger differences between treatments 3 and 4, and 7 and 8, and biases the differences the test statistic towards zero.

equity holdings) could have similar undesired effects. As before, we desire to be careful in making too bold conclusions. We merely point out that our evidence is consistent with such a role of complexity.

Adding further evidence to this, our separate experiment also finds that asset complexity is a factor in this under-reaction to incentives. Even university students, who compare favorably with the general population on proxies for cognitive ability, react more strongly to shifts in the return of an asset with a constant return than to shifts in an asset with a stochastic return when both shifts are economically equivalent. To our knowledge, this is a phenomenon that has not yet been documented in the literature on financial literacy, with the exception of the related effects in Chetty et al. (2009) and Abeler and Jäger (2015). This phenomenon raises questions for the psychology of arithmetic (Ashcraft, 1992) and has potentially numerous applications in the realm of economic decision making—think about changing incentive structures in deterministic vs stochastic environments. It raises also the general question to what degree a lack of understanding contributes to our results, in particular in the SOEP experiment. While it is hard to diagnose the presence of deep or full understanding of the choice task, it appears clear that even some basic understanding of the notion that earnings are tied to the return should lead to *some* response to our return manipulations. Consequently, the problem that we detect appears to relate fundamentally to the decision process and not only to its inputs although, as discussed in deatil in Appendix E, understanding of the environment does play some role for the rationality of choices.

For future research, our study may also inform the design of simple pilots for interventions regarding financial investment of households. In particular, in the light of the current underfunding of many pension systems (both pay as you go and capital funded), greater stock market participation by the middle class appears desirable to many economists and policy makers. Testing interventions in artefactual field experiments such as ours might avoid costly mistakes.

References

Abeler, J., & Jäger, S. (2015). Complex Tax Incentives. American Economic Journal: Economic Policy, 7(3), 1–28.

Allen, F. (1987). Discovering personal probabilities when utility functions are unknown. Management Science, 33(4), 542–544.

Andersen, S., Harrison, G., Lau, M., & Rutström, E. (2008). Eliciting Risk and Time Preferences. Econometrica, 76(3), 583–618.

Armantier, O., & Treich, N. (2013). Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging. European Economic Review, 62, 17–40.

Arrondel, L., Calvo-Pardo, H., & Tas, D. (2012). Subjective return expectations, information and stock market participation: Evidence from france. Retrieved from http://www.economics.soton.ac.uk/staff/calvo/documents/SubjectiveExpectations-16012012.pdf

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. Cognition, 44(1-2), 75–106.

Bellemare, C., Bissonnette, L., & Kröger, S. (2012). Flexible Approximation of Subjective Expectations Using Probability Questions. Journal of Business & Economic Statistics, 30(1), 125–131.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences. Journal of the European Economic Association, 11(6), 1231–1255.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. Quarterly Journal of Economics, 125(1), 263–306.

Binswanger, J., & Salm, M. (2013). Does everyone use probabilities? intuitive and rational decisions about stockholding (IZA Discussion Papers No. 7265). Institute for the Study of Labor (IZA).

Bucher-Koenen, T., & Lusardi, A. (2011). Financial Literacy and Retirement Planning in Germany (Working Paper No. 17110). National Bureau of Economic Research.

Campbell, J. Y., & Viceira, L. M. (2002). Strategic Asset Allocation: Portfolio Choice for Long-term Investors. New York: Oxford University Press.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and Taxation: Theory and Evidence. American Economic Review, 99(4), 1145–1177.

Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who Is (More) Rational? American Economic Review, 104(6), 1518–1550.

Costa-Gomes, M., Huck, S., & Weizsäcker, G. (2014). Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect. Games and Economic Behavior, 88, 298–309.

Delavande, A., Giné, X., & McKenzie, D. (2011). Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive are Answers to Variations in Elicitation Design? Journal of Applied Econometrics, 497, 479– 497.

Delavande, A., & Rohwedder, S. (2008). Eliciting Subjective Probabilities in Internet Surveys. Public Opinion Quarterly, 72(5), 866–891.

Deutsches Aktieninstitut. (2012). Zahl der Aktienbesitzer geht wieder zurück (Tech. Rep.). Frankfurt am Main: Deutsches Aktieninstitut. Retrieved from https://www.dai.de/files/dai{_}usercontent/dokumente/studien/2013-01{_}DAI-Kurzstudie{_}Aktionaerszahlen.pdf

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2013). Ambiguity Aversion and Household Portfolio Choice: Empirical Evidence. NBER Working Paper(18743).

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes : Measurement, Determinants and Behavioral Consequences. Journal of the European Economic Association, 9(3), 522–550.

Dominitz, J., & Manski, C. F. (2007). Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study. Journal of the European Economic Association, 5(2-3), 369–379.

Dominitz, J., & Manski, C. F. (2011). Measuring and Interpreting Expectations of Equity Returns. Journal of Applied Econometrics, 26, 352–370.

Forsythe, G. E., Malcolm, M. A., & Moler, C. B. (1977). Computer Methods for Mathematical Computations. Englewood Cliffs, NJ: Prentice-Hall.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. Proceedings of the National Academy of Sciences of the United States of America, 110(28), 11267–71.

Goldstein, D. G., & Rothschild, D. (2014). Lay Understanding of Probability Distributions. Judgment and Decision Making, 9(1), 1–14.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association, 1(1), 114–125.

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and Stock Market Participation. Journal of Finance, 66(6), 2121–2164.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. Journal of Finance, 63(6), 2557–2600.

Haliassos, M., & Bertaut, C. C. (1995). Why Do So Few Hold Stocks. Economic Journal, 105(September), 1110–1129.
Huck, S., & Müller, W. (2012). Allais for all: Revisiting the Paradox in a Large Representative Sample. Journal of Risk and Uncertainty, 44(3), 261–293.
Huck, S., & Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. Journal of Economic Psychology, 20, 699–715.
Hurd, M. D., & Rohwedder, S. (2012). Stock Price Expectations and Stock Trading. NBER Working Paper(17973).
Hurd, M. D., van Rooij, M., & Winter, J. (2011). Stock Market Expectations of Dutch Households. Journal of Applied Econometrics, 26, 416–436.
Hyman, J. (1983). Accurate Monotonicity Preserving Cubic Interpolation. SIAM Journal on Scientific and Statistical Computing, 4(4), 645–654.
Karni, E., & Safra, Z. (1995). The Impossibility of Experimental Elicitation of Subjective Probabilities. Theory and Decision, 38, 313–320.
Kézdi, G., & Willis, R. J. (2009). Stock Market Expectations and Portfolio Choice of American Households. Retrieved from `http://www.personal.ceu.hu/staff/Gabor{_}Kezdi/WorkingPapers/Kezdi-Willis-July2009-complete.pdf`
Lusardi, A., & Mitchell, O. S. (2011). Financial Literacy and Planning: Implications for Retirement Wellbeing (Working Paper No. 17078). National Bureau of Economic Research.
Merkle, C., & Weber, M. (2014). Do investors put their money where their mouth is? Stock market expectations and investing behavior. Journal of Banking & Finance, 46, 372–386.
Merton, R. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. Review of Economics and Statistics, 51(3), 247–257.
Naef, M., & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination (IZA Discussion Papers No. 4087). Institute for the Study of Labor (IZA).
Rabin, M., & Weizsäcker, G. (2009). Narrow Bracketing and Dominated Choices. American Economic Review, 99(4), 1508–1543.
Richter, D., & Schupp, J. (2012). SOEP Innovation Sample (SOEP-IS) — Description and Structure of the SOEP Innovation Sample. SOEPpapers(463). Retrieved from `https://www.diw.de/documents/publikationen/73/diw{_}01.c.407141.de/diw{_}sp0463.pdf`
Rothschild, D. (2012). Expectations: Point-estimates, probability

distributions, confidence, and forecasts. Retrieved from http://researchdmr.com/RothschildConfidence

Samuelson, P. (1969). Lifetime Portfolio Selection by Dynamic Stochastic Programming. Review of Economics and Statistics, 51(3), 239–246.

Stehle, R., Huber, R., & Maier, J. (1996). Rückberechnung des DAX für die Jahre 1955 bis 1987. Retrieved from http://edoc.hu-berlin.de/series/sfb-373-papers/1996-7/PDF/7.pdf

Stehle, R., Wulff, C., & Richter, Y. (1999). Die rendite deutscher aktien in der nachkriegszeit - dax-rückberechnung für die jahre 1948 bis 1955. Retrieved from http://www.econbiz.de/archiv/b/hub/bank/rendite{_}blue-chip-aktien.pdf

Trautmann, S. T., & van de Kuilen, G. (2015). Belief Elicitation: A Horse Race among Truth Serums. Economic Journal, 125(589), 2116–2135.

van Rooij, M., Lusardi, A., & Alessie, R. (2007). Financial Literacy and Stock Market Participation (Working Paper No. 13565). National Bureau of Economic Research.

Vissing-Jorgensen, A. (2004). Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth? Evidence from Expectations and Actions. NBER Macroeconomics Annual 2003, 18.

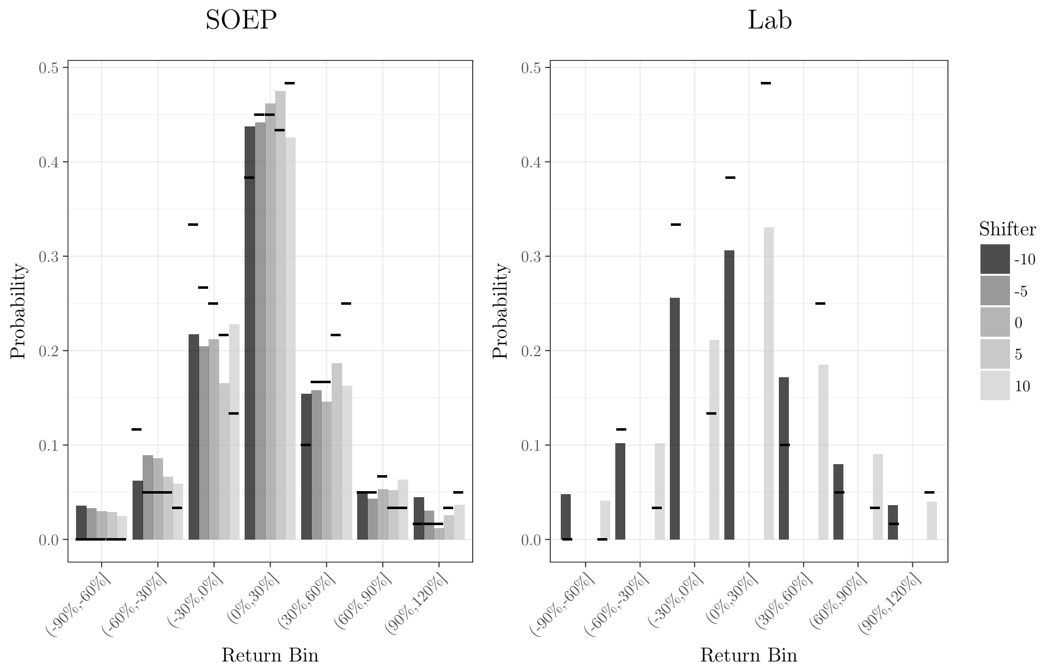
von Gaudecker, H.-M. (2015). How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice? Journal of Finance, 70(2), 489–507.

von Gaudecker, H. M., van Soest, A., & Wengström, E. (2011). Heterogeneity in Risky Choice Behavior in a Broad Population. American Economic Review, 101, 664–694.

Wilcox, N. T. (1993). Lottery Choice: Incentives, Complexity and Decision Time. Economic Journal, 103(421), 1397–1417.

Appendices

A Calibration



Historical benchmark for each treatment indicated by black horizontal lines.

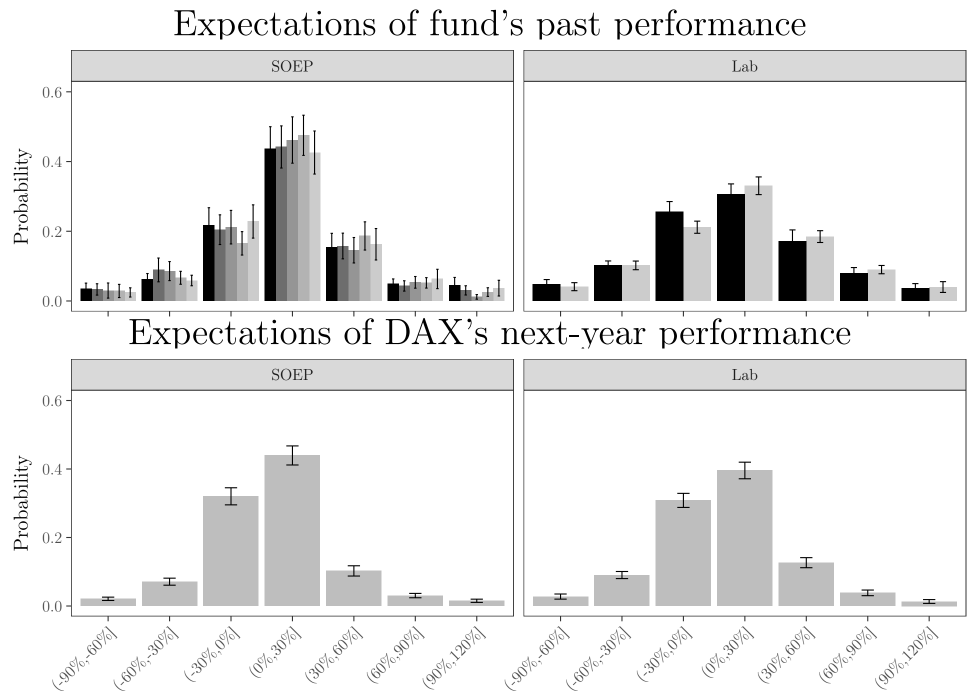
Figure A1: Historical distribution of returns vs. the average distributions in Lab and SOEP

Figure A1 compares the respondents' beliefs about the fund's return with the true historical distribution of DAX returns. The figure shows, in different shades of gray and ordered from left to right within each bin, the five different distributions of beliefs for the five different treatments. The figure also compares these distributions with five corresponding true distributions, indicated by black horizontal lines for each bin and treatment, that result from the true historical distribution plus the five shifters (in the same order, that is, from -10 to the very left to +10 to the very right, within each bin). The figure shows that SOEP respondents are remarkably well calibrated. In none of the seven bins are respondents off by more than 5 percentage points when data are pooled across treatments. The largest two deviations are that the frequency of small losses between 0 and 30% is slightly underestimated and the frequency of larger losses is slightly overestimated. The good calibration can also be seen in other metrics. While the mean return on the DAX from 1951 to 2010 was 15.5%, both the imputed and the stated expected return on the experimental asset of 12.5% and 8.3% respectively—while lower—are at least similar in magnitude to the historical mean. Moreover, while the relative frequency of a positive return over these six decades was 70.0%, SOEP respondents thought the DAX had seen a gain 69.3% of the time.[35] In contrast, the average distribution of our student subjects in the lab (also shown in Figure A1) differs significantly from the historical benchmark in that too much probability mass is assumed to be in the tails of the distribution.

Underneath the excellent calibration of the average SOEP respondent's belief lies, however, substantial heterogeneity in beliefs and miscalibration at the individual level. Very few of the distributions provided by individual respondents are close to the historical benchmark, and what produces the excellent calibration in the aggregate is a mixture of respondents who put the entire probability mass into a single bin and respondents who report diffuse distributions.

That the return expectations we elicit show such remarkable calibration stands in contrast to evidence from other countries, where substantial miscalibration is commonly observed. For the US Kézdi and Willis (2009) report that HRS respondents expected a stock market gain with roughly 50% probability in the 2002, 2004 and 2006 waves while the historical frequency of a gain on the Dow Jones was 68%. Similarly, the probability of a gain larger than 10% was estimated at 39% but the corresponding frequency was 49%. Dominitz and Manski (2011) find similar numbers in the monthly surveys of the Michigan

[35]In order to predict whether subjects invest in the risky asset, a relevant question—under expected utility, the only relevant question—is whether respondents expect a strictly positive excess return, i.e. a mean return that exceeds 4%. Based on reported beliefs, the proportion of respondents who expect a strictly positive excess return is 69.2% when using stated beliefs, and 72.6% when using imputed beliefs. The historical frequency of the DAX returning strictly more than 4% is 68.3%.



Error bars are 95% confidence interval.

Figure A2: Average distributions of past and future returns

Survey of Consumers from mid-2002 to mid-2004. In the Netherlands, Hurd et al. (2011) find that in 2004 the median expected rate of return on the Dutch stock market index was a mere 0.3%, a severe underestimate of the historical median return of 14%. A downward bias in expectations is by no means a universal finding, however. Respondents in the 1999, 2000 and 2001 waves of the Survey of Economic Expectations reported expectations for the S&P500 that were substantially above the historical average, but also held the S&P500 to be more volatile than has been the case historically (Dominitz & Manski, 2011).

What explains these differences with the existing literature? One possible explanation is that the papers quoted above compare respondents' expectations about the future with returns realized in the past. A test for correct calibration in this setting then amounts to a joint test of whether subjects hold the historical distribution of returns to be identical to the distribution of returns in the future and, if so, whether they have an accurate picture of the historical distribution. In contrast, we elicit beliefs about the distribution of returns over a well-defined period of time in the past and can test for calibration without auxiliary assumptions. The beliefs that we elicit about the next 12 months look, however, fairly similar, if somewhat more pessimistic – see Figure A2. This may not be entirely surprising as the survey period was just after the economic crises in parts of Europe had reached their peak intensity. In contrast to expectations about the past, where SOEP respondents and students differed substantially (with the former being more realistic), we find virtually identical expectations about the future between the two samples. The mean imputed return is 12.5% while the probability of a gain on the DAX is thought to be 58.8% on average. 51.8% of subjects state that they expect a return that is higher than 4%.

B Equity Share and Beliefs – Regressions

	Dependent Variable: Equity Share									
	SOEP: Stated Beliefs			SOEP: Imputed Beliefs			Lab: Stated Beliefs		Lab: Imputed Beliefs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Imputed Expected Return	0.003^{***}	0.005^{***}	0.005^{***}				0.003^{**}	0.007^{***}		
	(0.0005)	(0.001)	(0.001)				(0.001)	(0.003)		
Imputed Expected Return^2		-0.00002^{***}	-0.00001					-0.0001		
		(0.00001)	(0.00001)					(0.0001)		
Imputed Expected Return^3		-0.00000^{**}	-0.00000^{***}					-0.00000		
		(0.00000)	(0.00000)					(0.00000)		
Imputed S.D. of Return			0.001							
			(0.001)							
Probability of a Gain			-0.010							
			(0.037)							
Stated Expected Return				0.004^{***}	0.005^{***}	0.005^{***}			0.002	0.006
				(0.0005)	(0.001)	(0.001)			(0.002)	(0.005)
Stated Expected Return^2					-0.00001	-0.00001				-0.0001
					(0.00001)	(0.00001)				(0.0002)
Stated Expected Return^3					-0.00000	-0.00000				0.00000
					(0.00000)	(0.00000)				(0.00000)
Constant	0.330^{***}	0.330^{***}	0.370^{***}	0.340^{***}	0.330^{***}	0.400^{***}	0.420^{***}	0.410^{***}	0.440^{***}	0.420^{***}
	(0.012)	(0.013)	(0.110)	(0.011)	(0.013)	(0.110)	(0.028)	(0.035)	(0.030)	(0.037)
Personal Controls	No	No	No	No	No	No	No	No	No	No
N	562	562	560	562	562	560	198	198	198	198
R^2	0.074	0.093	0.160	0.081	0.090	0.140	0.031	0.063	0.016	0.038
Adjusted R^2	0.072	0.088	0.120	0.080	0.085	0.100	0.026	0.048	0.011	0.023

$^*p < .1$; $^{**}p < .05$; $^{***}p < .01$
Personal controls include dummy variables for gender, being born in the former GDR, having Abitur, having a university education, being employed, having a high self-assessed financial literacy, owning stocks and for each level of our wealth proxy. They also include age and age^2, household size, the number of children in the household and household income
All standard errors are Huber-White heteroskedasticity-robust

Table A1: Equity Share and Beliefs

C Different results for different people

In this section we exploit the rich data set on the SOEP respondents in order to study the role of socioeconomic background variables and direct measures or plausible correlates of savviness. As described in Section 6, we find strong differences between the SOEP sample and the university student sample regarding the extent to which they react to incentives. This raises the question of whether there is other evidence that “smart”, financially savvy respondents react more strongly to variations in incentives. The following analysis confirms the existence of such differences.

We caution that our examination of heterogeneity in the SOEP sample is a “fishing exercise”. However, its results are largely in line with what other studies have documented before, namely the fundamental role of cognitive ability for financial decisions making.

Table A3 documents treatment effects on choices and beliefs for different subgroups. It shows that there are small subsamples of the population that do react to incentives. For respondents with a university degree, the coefficients indicate an increase in equity share of one percentage point per one percentage point increase in return. Moving from the worst shifter of -10 to the best shifter of +10, the equity share is predicted to increase by 20 percentage points. This is similar to the effect we observe in the laboratory study with university students where the equity share increases by 33 percentage points. Hence, it appears that the main difference between SOEP and lab is driven by selection on educational covariates.

The results for respondents with different wealth levels are somewhat mixed. For reasons one can only speculate about, the strongest treatment effect is observed for those who withhold information on income from interest. There is also a notable composition effect between the two largest categories: respondents with low but positive levels of income from interest are predicted to increase their equity share by 14 percentage points when we move from the worst to the best shifter. Those without any interest earnings are estimated to exhibit a negative treatment effect.

Among the financial literacy questions we find a heterogeneous treatment

	Stock Market Participant			
	All	Abitur	University Degree	Financially Literate
Equity Share	0.200***	0.370**	0.480	0.230**
	(0.064)	(0.180)	(0.300)	(0.110)
Female	−0.029	−0.120	−0.230	−0.049
	(0.030)	(0.110)	(0.150)	(0.052)
Born in East Germany	−0.044	−0.021	−0.160	−0.083
	(0.033)	(0.120)	(0.190)	(0.061)
Age	0.004	−0.028	−0.062	0.002
	(0.006)	(0.023)	(0.044)	(0.011)
Age^2	−0.0001	0.0003	0.001	−0.00004
	(0.0001)	(0.0002)	(0.0005)	(0.0001)
Abitur	0.150**			0.240**
	(0.058)			(0.100)
University Degree	−0.003	−0.002		−0.041
	(0.072)	(0.097)		(0.120)
Household Size	−0.004	0.036	0.045	−0.020
	(0.022)	(0.087)	(0.110)	(0.035)
Risk Tolerance: Low	0.034	−0.015	−0.0003	0.048
	(0.035)	(0.110)	(0.140)	(0.059)
Risk Tolerance: High	0.058	−0.002	0.098	0.058
	(0.043)	(0.160)	(0.240)	(0.064)
Imputed expectation of DAX	0.0003	0.002	0.001	0.001
	(0.001)	(0.007)	(0.010)	(0.003)
S.D. of DAX	−0.001	−0.002	0.002	−0.001
	(0.001)	(0.004)	(0.005)	(0.002)
Gain Probability of DAX	0.039	−0.051	−0.330	0.062
	(0.085)	(0.310)	(0.480)	(0.160)
Number of Children in Household	−0.057*	−0.110	−0.180	−0.062
	(0.030)	(0.110)	(0.150)	(0.049)
Employed	−0.024	0.033	0.022	−0.007
	(0.037)	(0.120)	(0.210)	(0.067)
Financially Literate	0.080***	0.170*	0.200	
	(0.031)	(0.100)	(0.150)	
Interest: < 250 Euros	0.061*	0.047	−0.033	0.086
	(0.033)	(0.110)	(0.170)	(0.054)
Interest: 250 - 1.000 Euros	0.270***	0.330**	0.270	0.320***
	(0.057)	(0.140)	(0.220)	(0.084)
Interest: 1.000 - 2.500 Euros	0.430***	0.560***	0.560**	0.440***
	(0.086)	(0.180)	(0.240)	(0.110)
Interest: > 2.500 Euros	0.310***	0.150	0.013	0.560***
	(0.110)	(0.170)	(0.300)	(0.170)
Interest: refused to answer	0.150	0.350	0.046	0.260
	(0.100)	(0.250)	(0.360)	(0.170)
Household Income (missing=0)	0.023	0.039	0.029	0.010
	(0.018)	(0.040)	(0.059)	(0.029)
Household Income: missing	0.210**	0.150	0.520	0.140
	(0.084)	(0.330)	(0.560)	(0.130)
Constant	−0.130	0.580	1.400	−0.007
	(0.140)	(0.490)	(0.910)	(0.260)
N	560	122	72	283
R^2	0.280	0.360	0.480	0.320
Adjusted R^2	0.250	0.220	0.260	0.260

*p < .1; **p < .05; ***p < .01

Standard errors are Huber-White heteroskedasticity-robust. Household income is set to zero where missing (48 cases). Moreover, a dummy variable is added to the regression which is 1 for the observations with missing household income. "Financially Literate" is an indicator variable which is 1 whenever the respondent states that he/she is either "good" or "very good" with financial matters. For details on this and the other variables, see Online Appendix H.

Table A2: Stock market participation by subgroups effect only for the compound interest question. The other variables that might capture financial literacy do not show significant interactions with the experimental treatment. While the results on financial literacy and wealth are a bit patchy, overall a picture emerges that is familiar from the literature. Even relatively simple investment tasks as the one we have implemented here appear to be cognitively so complex that sensible responses to variations in parameters are shown only by skilled and sophisticated subjects.

An inspection of the two right-hand columns of Table A3 reveals that when it comes to belief manipulation no systematic patterns emerge. Only one of the interactions is statistically significantly different from zero, but only marginally so.

Given that we can identify some subgroups that react better to incentives, it is not far-fetched to presume that we might also be able to detect a stronger external validity of investment levels for these groups. With less noise in behavior inside and presumably outside the laboratory, the measured correlations between the experimental equity share and stock market participation may increase. Table A2 shows the regression-based conditional correlates of stock market participation, separately for different subgroups. Indeed it is the case that "smarter" subsamples show stronger external validity.

	Equity Share				Imputed Expectation of Fund				Stated Expectation of Fund			
	Mean		Treatment Effect		Mean		Treatment Effect		Mean		Treatment Effect	
Education												
< University Degree	0.373	(0.011)	0.000	(0.002)	12.646	(0.922)	0.107	(0.139)	8.649	(0.815)	0.076	(0.113)
University Degree	0.349	(0.033)	0.010**	(0.004)	11.426	(2.619)	0.325	(0.353)	5.586	(2.039)	-0.115	(0.300)
Interest from Wealth												
0	0.368	(0.017)	-0.005**	(0.002)	13.265	(1.572)	0.110	(0.224)	9.012	(1.597)	0.086	(0.214)
< 250 Euros	0.360	(0.019)	0.007***	(0.003)	10.576	(1.344)	0.320	(0.207)	7.759	(1.113)	0.076	(0.163)
250 - 1.000 Euros	0.344	(0.027)	0.001	(0.004)	18.231	(1.758)	-0.123	(0.297)	9.618	(1.569)	-0.247	(0.301)
1.000 - 2.500 Euros	0.422	(0.048)	-0.005	(0.007)	13.582	(3.266)	0.501	(0.518)	7.783	(1.846)	0.011	(0.204)
> 2.500 Euros	0.382	(0.054)	0.004	(0.007)	7.830	(8.722)	-0.653	(1.246)	5.481	(3.307)	0.206	(0.246)
refused to answer	0.339	(0.073)	0.015**	(0.007)	1.971	(8.978)	0.558	(1.030)	3.353	(3.572)	0.543	(0.351)
Financial Literacy: self-assessed												
'good' or 'very good'	0.360	(0.015)	0.002	(0.002)	14.064	(1.231)	0.287	(0.180)	8.047	(1.059)	0.153	(0.153)
'a little' or 'not at all'	0.381	(0.016)	-0.001	(0.002)	11.052	(1.227)	-0.001	(0.183)	8.479	(1.091)	-0.056	(0.147)
Financial Literacy: compound interest												
correct	0.384	(0.014)	0.004*	(0.002)	13.066	(1.157)	0.177	(0.178)	8.741	(0.865)	0.080	(0.117)
incorrect	0.349	(0.018)	-0.003	(0.003)	11.381	(1.415)	0.119	(0.190)	7.701	(1.431)	0.004	(0.213)
don't know	0.365	(0.059)	-0.003	(0.006)	15.608	(3.751)	-0.161	(0.547)	8.560	(4.725)	0.005	(0.533)
Financial Literacy: volatility												
correct	0.400	(0.047)	-0.005	(0.007)	21.056	(4.591)	-0.415	(0.664)	14.726	(4.607)	-0.763	(0.640)
incorrect	0.372	(0.012)	0.001	(0.002)	11.938	(0.906)	0.161	(0.134)	7.911	(0.755)	0.084	(0.102)
don't know	0.301	(0.041)	0.003	(0.006)	11.234	(3.342)	0.556	(0.439)	4.944	(3.744)	0.980*	(0.561)
Stock Owner												
yes	0.448	(0.028)	-0.002	(0.004)	12.828	(1.756)	-0.054	(0.308)	9.280	(1.417)	-0.439*	(0.237)
no	0.353	(0.011)	0.002	(0.002)	12.483	(0.992)	0.185	(0.142)	8.099	(0.878)	0.157	(0.118)

The table shows the results of multivariate regressions in which, for each set of rows, the outcome variables in the columns are regressed on indicator variables for the different levels of the row variables and a variable for the size of the shifter interacted with the different levels of the row variables. "Mean" and "Treatment Effect" therefore correspond to the constants and slope coefficients in bivariate regressions of the column variables on each of the different levels of the row variables. Standard errors for OLS regressions are Huber-White heteroskedasticity-robust.

Table A3: Treatment effect by subgroups

Table A4 describes (experimental) equity share of SOEP respondents and uses an indicator explanatory variable "interest", defined as one if the interest from wealth is positive and zero otherwise (after deletion of missing values). This indicator is interacted with the treatment variable. The coefficients show that introducing this interaction does not change the results for participants with university degree. In contrast, for participants without a university degree, we find that those without interest earnings have a negative reaction to the treatment while those with positive interest earnings show an insignificantly positive reaction.

	Equity share	
	SOEP with uni	SOEP without uni
	(1)	(2)
Treatment	0.011	-0.007^{***}
	(0.010)	(0.002)
Interest	0.130	-0.034
	(0.083)	(0.026)
Treatment*Interest	-0.0004	0.010^{***}
	(0.012)	(0.004)
Constant	0.260^{***}	0.380^{***}
	(0.066)	(0.017)
Observations	46	335
R^2	0.150	0.030
Adjusted R^2	0.089	0.021
Residual Std. Error	0.260 (df = 42)	0.240 (df = 331)
F Statistic	2.500^{*} (df = 3; 42)	3.400^{**} (df = 3; 331)

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Regression of Equity Share on Treatment Effect, Interest on wealth, and its interaction term. Results are presented for individuals from the SOEP with and without university degree.

Table A4: Treatment effect by subgroups with heterogeneity in interest

Table A5 shows the regression results for financial literacy (instead of "interest") interacted with the treatment variable. More precisely, we use an indicator "Financially Literate" that is one whenever the respondent states that he/she is either "good" or "very good" with financial matters and zero otherwise. In contrast to the results presented in Table A4, we can now compare SOEP respondents and lab participants as financial literacy is also available for the latter. From Table A5 we see that those with a combination of university degree and high financial literacy show the highest reaction to the treatment among the SOEP participants. Those without university degree show no effect, irrespective of their degree of financial literacy. Finally, the lab participants show the highest degree of reaction on treatment while the interaction term with financial literacy appears to be irrelevant.

	Equity share		
	SOEP with uni	SOEP without uni	Lab
	(1)	(2)	(3)
Treatment	0.005	−0.001	0.017***
	(0.008)	(0.002)	(0.002)
Financial Lit.	−0.013	−0.022	0.074
	(0.068)	(0.023)	(0.050)
Treatment*Financial Lit.	0.007	0.002	−0.001
	(0.010)	(0.003)	(0.005)
Constant	0.350***	0.380***	0.450***
	(0.050)	(0.017)	(0.021)
Observations	72	488	196
R^2	0.069	0.003	0.300
Adjusted R^2	0.028	−0.004	0.290
Residual Std. Error	0.280 (df = 68)	0.250 (df = 484)	0.260 (df = 192)
F Statistic	1.700 (df = 3; 68)	0.430 (df = 3; 484)	28.000*** (df = 3; 192)

*p<0.1; **p<0.05; ***p<0.01

Regression of Equity Share on Treatment Effect, financial literacy, and its interaction term. Results are presented for individuals from the SOEP with and without university degree and individuals from the laboratory sample.

Table A5: Treatment effect by subgroups with heterogeneity in financial literacy

In Table A6, we depict the regression results for stock ownership interacted again with the treatment variable. For the SOEP participants with university degree we see that the interaction term is statistically irrelevant.

	Equity share	
	SOEP with uni	SOEP without uni
	(1)	(2)
Treatment	0.009**	0.001
	(0.005)	(0.002)
StockOwnership	0.170**	0.087**
	(0.073)	(0.034)
Treatment*StockOwnership	0.001	−0.007
	(0.010)	(0.005)
Constant	0.290***	0.360***
	(0.036)	(0.012)
Observations	72	489
R^2	0.140	0.020
Adjusted R^2	0.110	0.014
Residual Std. Error	0.270 (df = 68)	0.250 (df = 485)
F Statistic	3.800** (df = 3; 68)	3.300** (df = 3; 485)

*p<0.1; **p<0.05; ***p<0.01

Regression of Equity Share on Treatment Effect, Stock ownership, and its interaction term. Results are presented for individuals from the SOEP with and without university degree.

Table A6: Treatment effect by subgroups with heterogeneity in stock ownership

D Measurement Error

As a proxy variable for measurement error we compute the difference between actual and imputed beliefs and rank individuals according to the absolute value of it. We emphasize that the ranking is performed for SOEP and Lab participants separately. In the following analysis we compare two different cases.

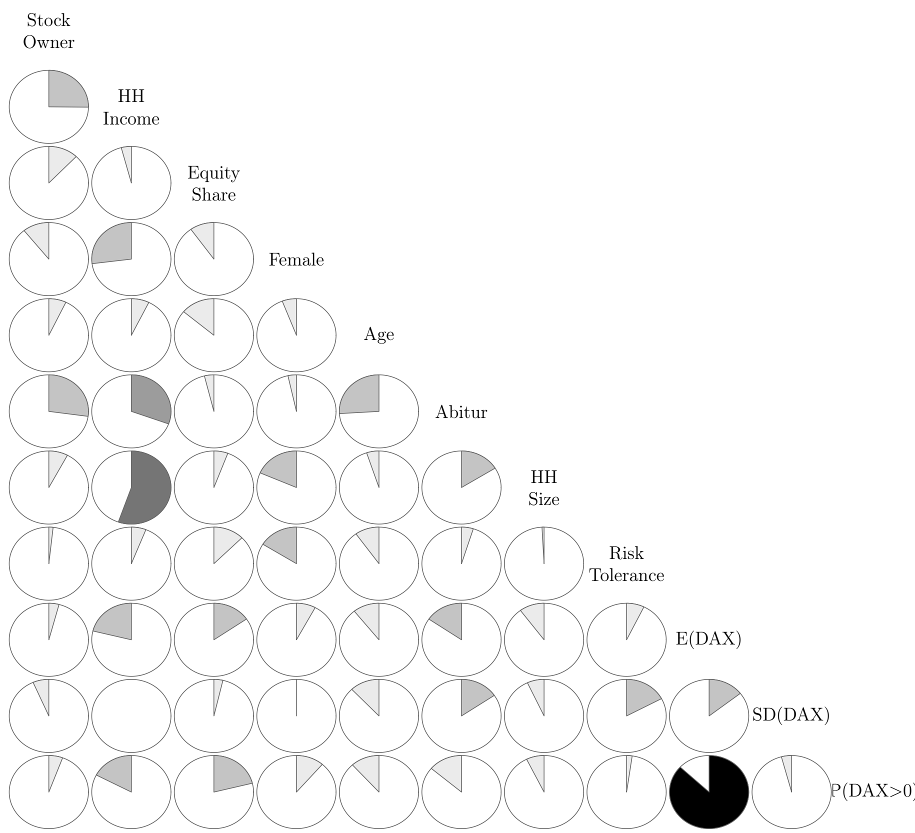


Figure A3: Correlogram as in Figure 2 but with "low measurement error"–individuals only.

First, we consider the case of *low measurement error*, i.e., we only include individuals with a measurement error proxy variable below the first quartile. Second, we analyze the *high measurement error* case, which contains those individuals only with a measurement error proxy variable above the third quartile.

Figure A3 depicts a correlogram for individuals with low measurement error. We generated the correlogram with back of the envelope calculations reported in Section 5 for the subset of respondents whose measurement error in beliefs is small. The results show mild but insignificant changes in the expected directions.

	Equity share			
	SOEP low ME	Lab low ME	SOEP high ME	Lab high ME
Treatment	0.004	0.009**	0.002	0.016***
	(0.003)	(0.004)	(0.003)	(0.004)
Constant	0.360***	0.480***	0.340***	0.450***
	(0.020)	(0.042)	(0.020)	(0.040)
Observations	162	50	142	50
R^2	0.009	0.098	0.003	0.280
Adjusted R^2	0.003	0.079	−0.004	0.260
Res. Std. Err.	0.250 (df=160)	0.290 (df = 48)	0.240 (df=140)	0.260 (df = 48)
F Statistic	1.50 (df=1;160)	5.20** (df=1;48)	0.49 (df=1;140)	18.00*** (df=1;48)

*p<0.1; **p<0.05; ***p<0.01

Table A7: Regression of equity share on treatment for different subgroups of measurement error (ME), i.e., ordered absolute distance between actual and imputed beliefs is below first (low) or above third (high) quartile.

For different subgroups of ranked individuals, Table A7 depicts the estimated effect of treatment on the equity share. From this table we see that for SOEP participants we have positive but not significant effects.

We further investigate the measurement error issue by analyzing alternative ways to account for it. The following tables show results when measurement error is estimated via the absolute difference between the mean of historical DAX returns and the mean of imputed and stated beliefs. In the following, we refer to individuals with low measurement error this ordered absolute distance is below the first quartile and to individuals with high measurement error if it is above the third quartile.

Table A8 depicts the estimation results for regressing equity share on treatment for different measurement error groups. Again we see that for SOEP individuals the treatment effect is positive but not significant.

	Equity share			
	SOEP low ME	Lab low ME	SOEP high ME	Lab high ME
Treatment	0.004	0.008**	0.002	0.015***
	(0.003)	(0.004)	(0.003)	(0.005)
Constant	0.420***	0.550***	0.270***	0.360***
	(0.020)	(0.037)	(0.021)	(0.048)
Observations	142	50	141	51
R^2	0.014	0.094	0.005	0.220
Adjusted R^2	0.003	0.079	−0.003	0.200
Res. Std. Err.	0.240 (df = 140)	0.240 (df = 48)	0.240 (df = 139)	0.270 (df = 49)
F Statistic	2.00(df=1;140)	5.00**(df=1;48)	0.64(df=1;=139)	14.00***(df=1;49)

*p<0.1; **p<0.05; ***p<0.01

Table A8: Regression of equity share on treatment for different subgroups of measurement error (ME), i.e., ordered absolute distance between mean of historical DAX returns and mean of imputed and stated beliefs is below first (low) or above third (high) quartile.

E Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Female	700	0.480	0.500	0	1
Age	700	53.000	17.000	16	94
Born in Germany	700	0.860	0.350	0	1
Born in the GDR	700	0.200	0.400	0	1
Abitur	700	0.200	0.400	0	1
University degree	700	0.120	0.330	0	1
Employed	700	0.500	0.500	0	1
Household Size	700	2.300	1.200	1	8
Number of Children in Household	700	0.360	0.780	0	6
Monthly Household Income (in 1000s of Euros)	652	2.500	1.500	0.100	12.000
Risk Tolerance	700	4.900	2.500	0	10
Financial Literacy (self-assessed: 'good' or 'very good')	697	0.500	0.500	0	1
Financial Literacy (compound interest question correct)	690	0.580	0.490	0	1
Financial Literacy (volatility question correct)	690	0.840	0.370	0	1
Equity share (in experiment)	562	0.370	0.260	0.000	1.000
Imputed expectation of fund	562	13.000	21.000	−80.000	110.000
Stated expectation of fund	562	8.300	18.000	−80.000	95.000
Gain Probability of Fund	562	0.690	0.280	0.000	1.000
Imputed expectation of DAX	562	5.500	18.000	−60.000	90.000
Gain Probability of DAX	562	0.590	0.330	0.000	1.000
Total Liquid Assets	515	19.000	44.000	0.000	446.000
Stock Market Participation	693	0.180	0.390	0	1
Stocks (amount)	671	1,780.000	7,874.000	0	110,000
Stocks / Total Liquid Assets	452	0.066	0.190	0.000	1.000
Total Debt	666	17,174.000	54,514.000	0	800,000

N is the number of non-missing observations

Table A9: Descriptive statistics for the 700 heads of household in SOEP sample

F Imputation of Moments

To derive various summary statistics from the elicited belief distributions we fit continuous distributions to the raw data and calculate the statistics from these distributions.

While much of the existing literature fits parametric distributions we follow an approach similar to Bellemare et al. (2012) and fit cubic interpolating splines using an approach due to Forsythe, Malcolm, and Moler (1977). We first cumulate the probabilities that respondents place within each of the seven bins. This yields 8 points on the cumulative distribution function from which the responses were generated. We take these 8 points to be the knots of the spline (that is, we ignore any rounding in the response and assume that the CDF at these points is known) and interpolate between them with a piecewise cubic polynomial.

Since each of the 7 pieces is defined by four polynomial coefficients this is a problem with 28 unknowns. The condition that the spline must go through each of the 8 points gives 14 equations (one each for the end-points and two each for the interior knots) and further assuming that the spline is twice continuously differentiable at each of the knots yields 12 additional equations. What pins down the spline are two boundary conditions, which are found by fitting exact cubics through the four points closest to each boundary and imposing the third derivatives of these cubics at the end-points on the spline.

What is problematic about using such a spline to impute a CDF is that nothing in the procedure described above guarantees that the resulting spline is monotonic. To overcome this problem we apply a filter to the spline that is due to Hyman (1983). The filter relaxes some of the smoothness conditions enough to ensure monotonicity.[36]

Figure A4 demonstrates the fit for six representative respondents. Circles show the raw cumulative probabilities to which both the Hyman-filtered cubic splines as well as various alternative distributions are fitted. By construction the splines are extremely close to the data in all cases – often much closer than

[36] Both the Forsythe et al. construction of the spline as well as the Hyman filter are implemented in `R` through the `splinefun()` function with `methods` `fmm` and `hyman` respectively

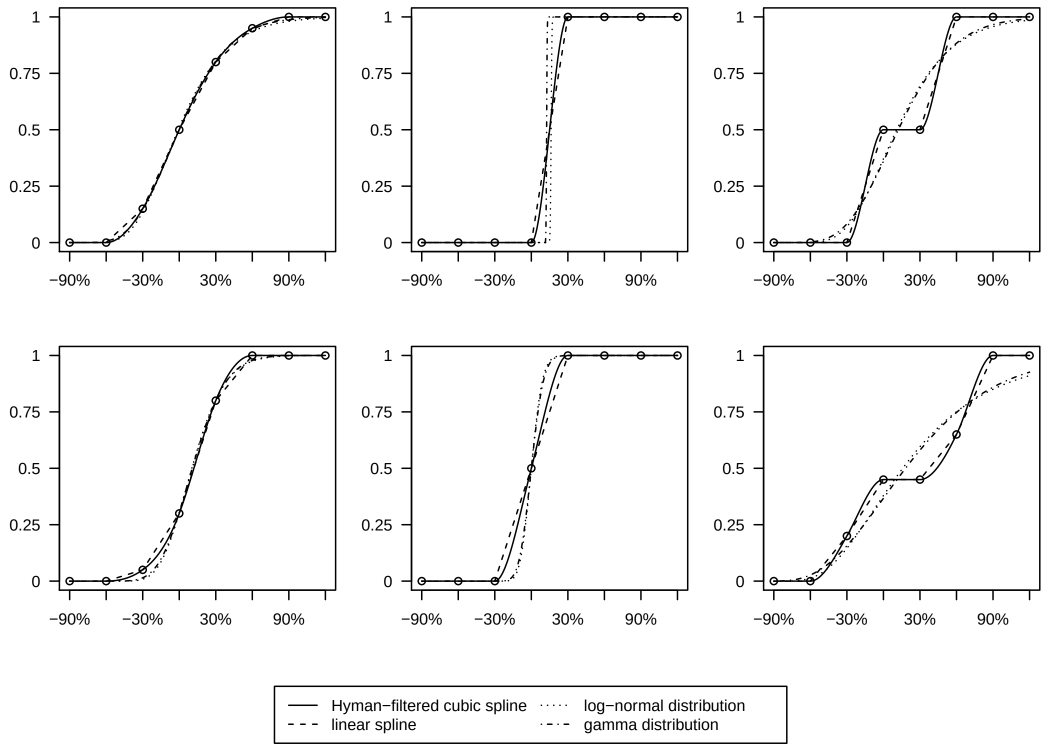


Figure A4: CDFs derived from the belief data using both spline interpolation and parametric distributions fit via least squares

any of the parametric distributions that have been fit to the data by minimizing the sum of squared deviations at the 8 points. The two distributions on the left are single-peaked and have non-zero probability in several bins and for these cases all of the methods yield roughly the same fit. The distributions in the middle have mass only in a single or in two of the bins, which is a problem for the parametric distributions because in such cases the fit can be improved ad infinitum by reducing the variance of the distribution and thereby reducing the sum of squared deviations at the 8 points. In the two cases on the right the distribution is multi-modal, which naturally leads to terrible fit for the parametric distributions, all of which are unimodal. The splines, in contrast make no such assumptions and therefore fit even these cases rather well.

Finally, we calculate both the mean and the standard deviation from these distributions numerically using adaptive Gauss-Kronrod quadrature.

G Some Individual Belief Distributions

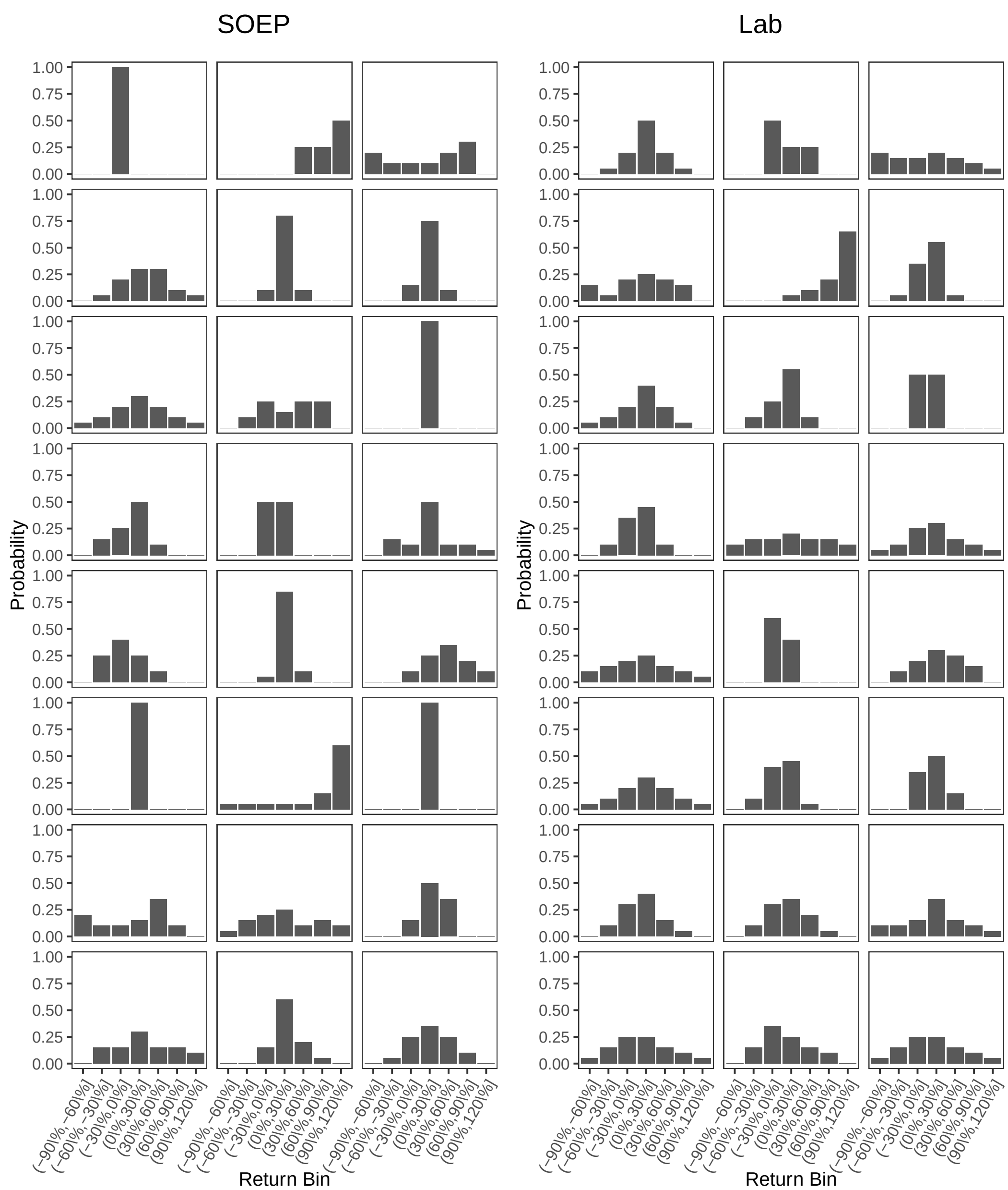


Figure A5: 24 randomly chosen belief distributions from both the SOEP and the lab sample.