The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

The Standard Portfolio Choice Problem in Germany

Steffen Huck
Tobias Schmidt
Georg Weizsäcker

CESIFO WORKING PAPER NO. 5441
CATEGORY 13: BEHAVIOURAL ECONOMICS
JULY 2015

An electronic version of the paper may be downloaded
- *from the SSRN website:* *www.SSRN.com*
- *from the RePEc website:* *www.RePEc.org*
- *from the CESifo website:* *www.CESifo-group.org/wp*

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5441*

The Standard Portfolio Choice Problem in Germany

Abstract

We study an investment experiment conducted 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: they predict real-life stock market participation. But many households do not significantly react 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 than modifying the risky asset's return.

JEL-Code: D100, D140, D840, G110.

Keywords: stock market expectations, stock market participation, portfolio choice, artefactual field experiment, financial literacy, complexity.

Steffen Huck
WZB Berlin
Reichpietschufer 50
Germany – 10785 Berlin
steffen.huck@wzb.eu

Tobias Schmidt
DIW Berlin
Mohrenstr. 58
Germany – 10117 Berlin
tschmidt@diw.de

Georg Weizsäcker
Humboldt University Berlin
Spandauer Str. 1
Germany – 10178 Berlin
weizsaecker@hu-berlin.de

July 20, 2015
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, 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 WZBTU decision laboratory in Berlin for their excellent contributions in the preparation and conduct of the experiments. Financial support by the ERC (Starting Grant 263412) is gratefully acknowledged.

1 Introduction

We report on an artefactual field experiment that examines investment behavior in a representative sample of the German population. The experiment uses households from the Socio-Economic Panel's "Innovation Sample" (SOEP-IS) as respondents. They act as investors who face a standard portfolio choice problem, allocating a fixed budget between a safe and a risky asset. No other investments are possible and the investment horizon is fixed. Despite its drastic simplification, the standard portfolio choice problem is widely viewed as capturing one of the main tradeoff in financial decision making. We regard its relevance as an empirical question and examine both its internal consistency and external validity for the German general population. Regarding external validity, behavior in our artefactual investment task is robustly correlated with actual stock market participation, even after controlling for many of the correlates of participation that the existing literature has identified. The average stock market particpation rate is 18% in our representative sample of households; and a one-standard-deviation increase in the experiment's investment in the risky asset is associated with an increase in stock market participation by 6 percentage points. Regarding internal consistency, we find that investments in the risky asset are correlated with measures of beliefs about the asset's return, lending further credibility to the story that the standard portfolio choice model sets out to tell. However, the data also shows how severely respondents' cognitive limitations and financial skills affect decisions. We exogenously vary the returns of the risky asset across treatment groups, by paying some groups a fixed percentage over and above the stock market return and some groups a fixed percentage below the stock market return, and find that only a subsample of relatively well-educated respondents reacts to such changes in incentives. For all other respondents, the opportunity to earn additional money is lost.

Alongside the artefactual field experiment, we also present a laboratory study in which we use the same protocol on a convenience sample of university students. The results are largely congruent between the two settings, with one notable difference: unlike the general population, university students *do* react to the variation in incentives.[1]

The above evidence points at an important role of task complexity for financial decision making. It is plausible that university students understand the incentive structure better and that this induces them, but not the typical German household, to react to the incentive change. That is, the perceived complexity of the incentive change may be higher for some people than for others. We also examine another channel. The two assets in the standard portfolio choice problem differ in nature, one being characterized by just a single number, the other by a (subjective) probability distribution. An investor may find it easier to appreciate a shift in the single number than in the probability distribution.

We test this new hypothesis in an additional laboratory experiment where economically equivalent incentive shifts come in two guises—once as a shift in the return of the risky asset and once as a shift in the safe return. The experimental design ensures that both incentive variations are equally easy to understand[2] and each participant faces both kinds of manipulations. The experimental results confirm that the reaction to changes in the safe asset is indeed significantly stronger than the reaction to changes in the risky asset. This pattern has not yet been observed in the literature, to our knowledge, and cannot be explained by standard theories of decision-making under uncertainty.[3] It has, potentially, important consequences for the optimal design of tax incentives for investments and other regulatory measures.

Relation to existing literature. Our experimental design builds on the sizable literatures on stock market participation, belief elicitation and experiments on choice under uncertainty. Our results are mostly, but not in all cases, consistent with these literatures and we emphasize some of the relevant comparisons.

[1] Notice, however, that the students, just as the SOEP participants, exhibit too mild a change in beliefs in response to incentive changes.

[2] The experiments involve incentive shifts for both assets, presented in the same format. A controlled variation of the shift sizes and a simultaneous variation of an illiquid asset generates the isomorphy within pairs of incentive shifts.

[3] One possible way of rationalizing the pattern is to posit that the manipulation of the assets affects the perceived source of uncertainty (in the sense of Fox and Tversky (1995), and Abdellaoui, Baillon, Placido, and Wakker (2010)).

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[4] and observational or experimental findings on financial literacy and subjective expectations abound (for survey evidence on financial literacy and its correlates in the German population, see Bucher-Koenen & Lusardi, 2011).

A growing literature measures the subjective beliefs of the general public about stock returns. The earliest survey questions asked for a measure of central tendency only (Vissing-Jorgensen, 2004). Questions to elicit entire distributions have more recently been added to many surveys.[5] These questions ask for statements about the probabilities of the market returns lying above given thresholds.[6] The broad picture emerging from this literature is that expectations are extremely heterogeneous, often lie far away from actual returns (Hurd et al., 2011)[7] and show positive predictive power for stock market investments.

One drawback of these methods is that responses are often internally in-

[4]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.

[5]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).

[6]In the Health and Retirement Survey respondents are asked for the "chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today" and the "chance they will have grown by 10 percent or more" (Dominitz & Manski, 2007). Assuming no measurement error these two questions yield two points on the CDF and, if one is willing to make distributional assumptions, allow fitting an entire distribution for every individual.

[7]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%.

consistent (Binswanger & Salm, 2013).[8] Instead of asking for probabilities of a return lying above a threshold, we use a histogram elicitation method pioneered by Delavande and Rohwedder (2008) in which respondents are asked to distribute a fixed number of items that jointly represent a probability mass of 1 into a number of bins. The method allows using all available data instead of focusing on consistent sets of responses. The method also has the advantage of being easy to understand; it has been successfully used even with respondents with little formal education and low numerical and statistical skills (Delavande, Giné, & McKenzie, 2011).[9]

In contrast to previous findings, the respondent in our sample report beliefs that accurately capture the historical market return distribution, at least in the aggregate. This is detailed in 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 our comparatively small sample size as well as to the different parts of the sample which enter into the econometric analysis (previous studies often discard the sizeable numbers of respondents who report internally contradictory beliefs). 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.[10]

While there is a large literature on how people make risky choices[11] and

[8]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 outright.

[9]We additionally ask respondents for a simple numerical expectation, which yields very similar results in most parts of the analysis.

[10]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.

[11]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).

on the relevant correlates[12], 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 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 when the after-tax price of goods does not change and react more weakly to changes in taxes that are applied at the register instead of being posted on the price tag. Abeler and Jäger (2014) 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, which is described by 30 rules. 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.[13]

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

[13] 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, elsewhere, investors are found to react to extraneous information such as advertisements for standard financial assets (like individual stocks) or photos of financial advisors (Bertrand, Karlan, Mullainathan, Shafir, & Zinman, 2010). This is also consistent with the findings of Binswanger and Salm (2013) who argue

The remainder of the paper is organized as follows. In Section 2 we describe the experimental design and procedures for both the household panel and the laboratory. In Section 3 we focus on the experimental data and study the relation between beliefs about returns and investments in the experiment. In Section 4 we turn to the validity questions that relate the experimental data to socioeconomic data from the household panel, and in Section 5 we examine the treatment effects. Section 6 presents the additional experiment comparing the return manipulation between safe and risky assets, and Section 7 concludes.

2 Experimental Design and Procedures

2.1 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 trial new questions and modules (see Richter & Schupp, 2012, for details). Its sampling of households follows the same procedure as the SOEP does and renders the SOEP-IS 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 long 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.

The 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.

that large subsamples of the population may not think probabilistically about stock market returns at all.

The core component of the module is the interactive experiment modelled on the standard portfolio choice problem that we describe in the following.[14][15]

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 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 should make 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

[14] 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.

[15] A complete set of instructions are available in the Supplementary Material.

possible there (see Section 2.2 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.[16]

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 €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.[17]

[16]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.

[17]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.

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 et al. (2011) and Rothschild (2012).[18] A screenshot of the interface can be found in Appendix D. Respondents have to place 20 "bricks", each representing a probability mass of 5%, into seven bins of possible percentage returns. 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[19] 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 incentivize either of these belief measures. Properly incentivizing subjects 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.[20]

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.

[18]For an overview of studies which have used this or similar methods see Goldstein and Rothschild (2014) and references therein.

[19]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.

[20]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 incentivising the elicitation of beliefs does not yield biased answers in these studies but merely noisier answers. A further concern with incentives is the introdction of possible motives for attempted hedging between tasks (see e.g. Karni & Safra, 1995).

	Dependent variable: Participation in the Experiment
Female	−0.001 (0.030)
Born in the GDR	0.028 (0.038)
Abitur	0.043 (0.058)
University Degree	−0.001 (0.070)
Household Size	−0.018 (0.019)
Number of Children in Household	0.019 (0.034)
Employed	0.017 (0.038)
Financially Literate	0.028 (0.030)
Interest: < 250 Euros	−0.028 (0.035)
Interest: 250 - 1.000 Euros	0.027 (0.049)
Interest: 1.000 - 2.500 Euros	0.096 (0.093)
Interest: > 2.500 Euros	0.120 (0.240)
Interest: refused to answer	−0.076 (0.087)
Stock Market Participant	0.025 (0.046)
Risk Tolerance: Low	0.029 (0.033)
Risk Tolerance: High	0.027 (0.041)
Age bracket 31-40	0.032 (0.077)
Age bracket 41-50	−0.083 (0.059)
Age bracket 51-60	−0.084 (0.057)
Age bracket 61-70	−0.064 (0.060)
Age bracket > 70	−0.200*** (0.059)
N	692

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

Table 1: Selection into the experiment: Probit marginal effects

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

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 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 sigificance of all variables other than the age brackets cannot reject the null of no effect ($\chi^2(18) = 19.41$, $p = 0.37$).

2.2 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.[21] 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

[21] The decison laboratory uses ORSEE (Greiner, 2015).

between survey and lab environments.

3 Experimental Data

3.1 Beliefs and Investments

We start with a summary description of investments and elicited beliefs. We will call the share of wealth 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.

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 representative 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 will often use summary statistics that we compute from the reported histograms. To compute statistics like the expectation or the standard deviation of returns from the underlying belief distribution 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

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

[23]Compared with other belief elicitations using similar methods these frequencies are on the low side. Delavande and Rohwedder (2008) report that 73% of their subjects used two or fewer bins.

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

	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

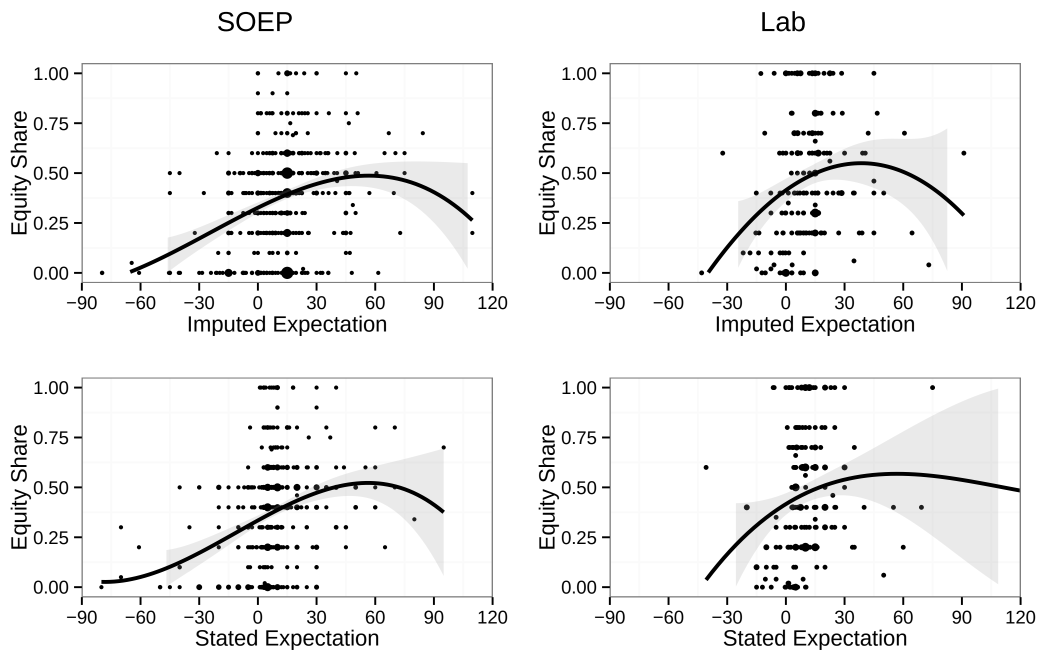
"Financially Literate" is an indicator variable which is 1 whenever the respondent states that he/she is either "good" or "very good" with financial matters. For details on this and the other variables, see Appendix H.

Table 2: Experimental Responses in the SOEP by subgroup

distribution is 24.0%. 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 in both settings (Pearson correlation coefficient: 0.5 for the SOEP and 0.31 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).

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 incrase in the equity share (see Figure 1 for OLS regressions). This relationship holds for both our belief measures and is roughly the same in the laboratory. This evidence of a 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).



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 A3 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 there are also 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 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.

4 External validity: Stock market participation

We now turn to the important question whether our response variables are indicative of real-life investments. Specifically, we test the external validity of our data by comparing elicited behavior in the experiment with survey responses to the question "Do you own any stock market mutual funds, stocks

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

or reverse convertible bonds ("Aktienanleihen")?"

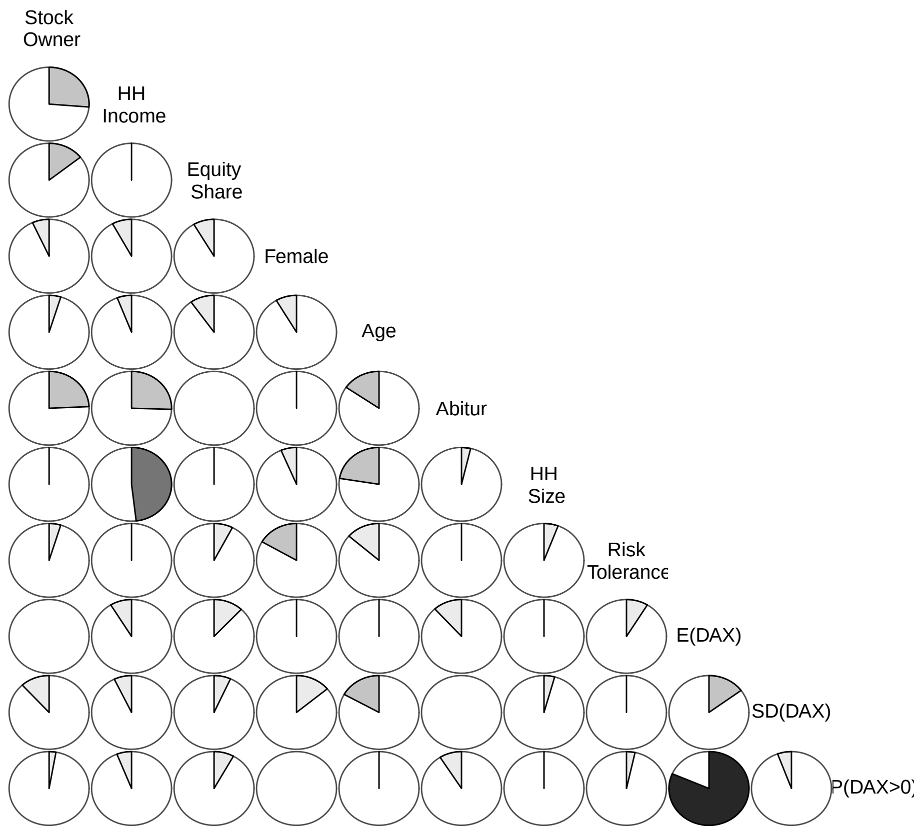
18% of all households answered this question in the affirmative, which is in line with other evidence on the German stock market participation.[25] Splitting the participation rate by deciles of both household income and a proxy for liquid wealth[26] Table 3 also shows that stock market participation 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 predictors 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 correlate 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 predictors of stock ownership. Notice that equity share is the only experimental variable that has predictive power for stock holdings (correlation: 0.14, p-value: < 0.001).

[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 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

Of course, the correlograms only show bivariate relations. In order to gain a broader picture we investigate whether the correlations change if we take into account other explanatory variables.[27] 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

[27]This is similar to the approach taken by Guiso et al. (2008) who study the co-variation of stock market participation with generalized trust and other variables.

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.

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 wind tunnel: 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.

	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

5 Treatment effects

Recall that we implement five exogenous treatments that shift the historical return of the DAX. The shifts are sizeable, ranging from -10 percentage points to +10 percentage points. Table 5 documents that by and large there is, surprisingly, no effect of the return shifter on equity share in the SOEP sample. 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 incorprate the shift appropriately in their investment choice.

However, this result is not universal. Instead we notice 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.[28]

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 whether we consider the SOEP data or the laboratory data and no

[28]For details, see Section C in the Appendix.

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

matter how we slice the 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.

We tentatively conclude from this evidence that it is much harder to manipulate beliefs than to elicit them. As we show in Appendix A subjects' 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 just like the investments, the beliefs do not react strongly enough to the experimental manipulation. This also raises the question how well the respondents understand the manipulation. The next section investigates the possibility that the weak reaction to the manipuation may be driven by factors beyond the understanding of the experimental instructions.

6 Asset Complexity and Reactions to Changes in Incentives

In this section we investigate the role of complexity with an additional laboratory experiment. We introduce manipulations of both, the risky and the safe, asset that are economically equivalent and described in identical terms. Yet, the experiment shows that the reaction to an increase in the stochastic return is weaker than the reaction to an increase in the fixed return. This effect is largely consistent with the available evidence on reactions to tax incentives as, e.g., in Chetty et al. (2009) and Abeler and Jäger (2014).

In the additional experiment, the excess return of the risky asset is varied in two ways: either a shift of Δ in the risky asset's return, or a shift of $-\Delta$ in the safe asset's return. To make the two shifts fully economically equivalent, we modify the decision maker's exogenous income level, as detailed in the next subsection. However, before proceeding to the details, two remarks are in order: First, we designed this section's experiment after we observed the results from the experiments described in Section 2.2—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, of course, but have no proof, that these subsamples would exhibit even larger differences in their reactions to different shifts.

6.1 Experimental Design

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 payment in addition to the earnings from the portfolio choice.

The participants are endowed with W_I worth of an illiquid asset that generates a sure income and 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 W_I by amount ΔW_L, yielding the same payoff from investing a share α in the risky asset:

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

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 about the risky asset into outcomes, 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.[29] 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

[29] 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

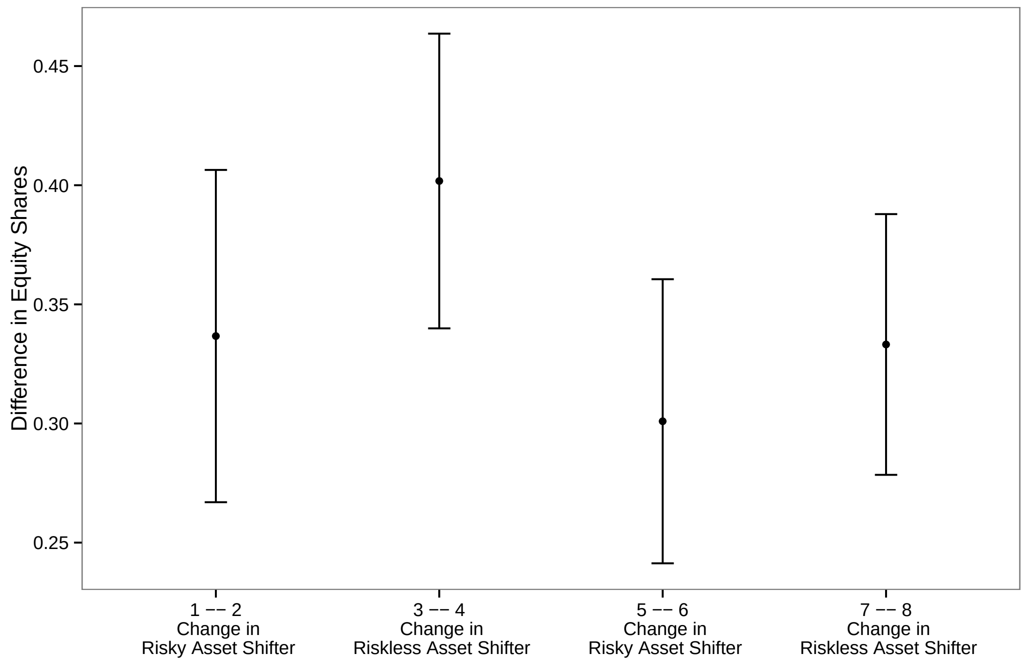
Table 6: Treatment parameters

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. The hypothesis of a stronger reaction tho shift in safe return suggests that investments 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 using identical procedures as in the study described in Section 2.2. 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.

6.2 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 pairs.



Error bars show 95% confidence intervals.

Figure 3: Investments in the risky asset by treatment

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 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.[30]

Given the comparatively small sample size, each of these mean responses is subject to considerable sampling error. In order to formally test our main

[30]A graph of the raw responses is available in Appendix J.

hypothesis we therefore pool the data from all treatments. We compute the difference in differences for treatments 1 to 4 and add to this the difference in differences for treatments 5 to 8. Under the null of rational, 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 (two-sided Wilcoxon rank sum test p-value = 0.03, two-sided t-test p-value = 0.09).[31]

7 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 representative population sample in Germany. The data from this experiment exhibit high degrees of external validity as shown through direct comparison of behavior inside and outside the experiment. This may be viewed as a success for the standard portfolio choice model. Despite its extreme reductionism it captures important real-life tradeoffs in financial markets.

The analysis also shows that the degree of external validity 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 financial decision problems. 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 forclosure 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

[31] 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.

for equity holdings) could have similar undesired effects.

In a separate experiment, we also find evidence 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 (2014). This phenomenon—that how well addition can be performed may depend on the nature of the variable to which a number is added–raises some new and deep issues for the psychology of arithmetic (Ashcraft, 1992) and has potentially numerous relevant applications for economic decision making in uncertain environments.

For future research, our study may inform the design of further wind tunnels 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

Abdellaoui, M., Baillon, A., Placido, L. t. L., & Wakker, P. P. (2010). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review*, *101*(2), 695–723.

Abeler, J., & Jäger, S. (2014). *Complex Tax Incentives.* Retrieved from `http://scholar.harvard.edu/files/jaeger/files/abelerjaeger2013_complex_tax_incentives.pdf`

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.

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`

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.

Fox, C. R., & Tversky, A. (1995). Ambiguity Aversion and Comparative Ignorance. *Quarterly Journal of Economics*, *45*(1), 97–113.

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. *Eco-*

nomic 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*.

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*.

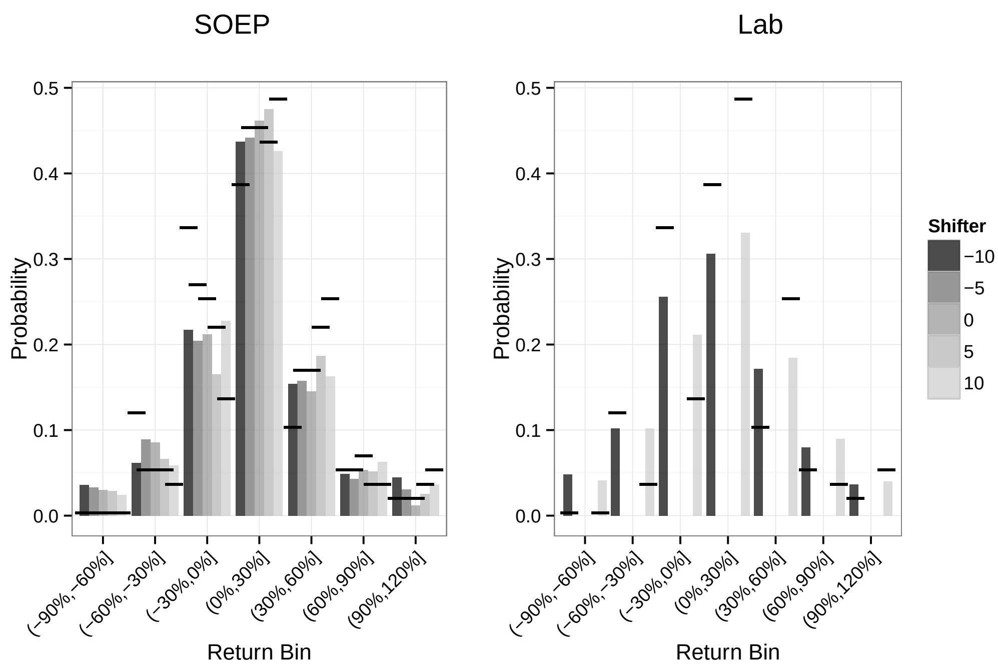
von Gaudecker, H.-M. (2015). How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice? *Journal of Finance*, *70*(2), 489–507.

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.

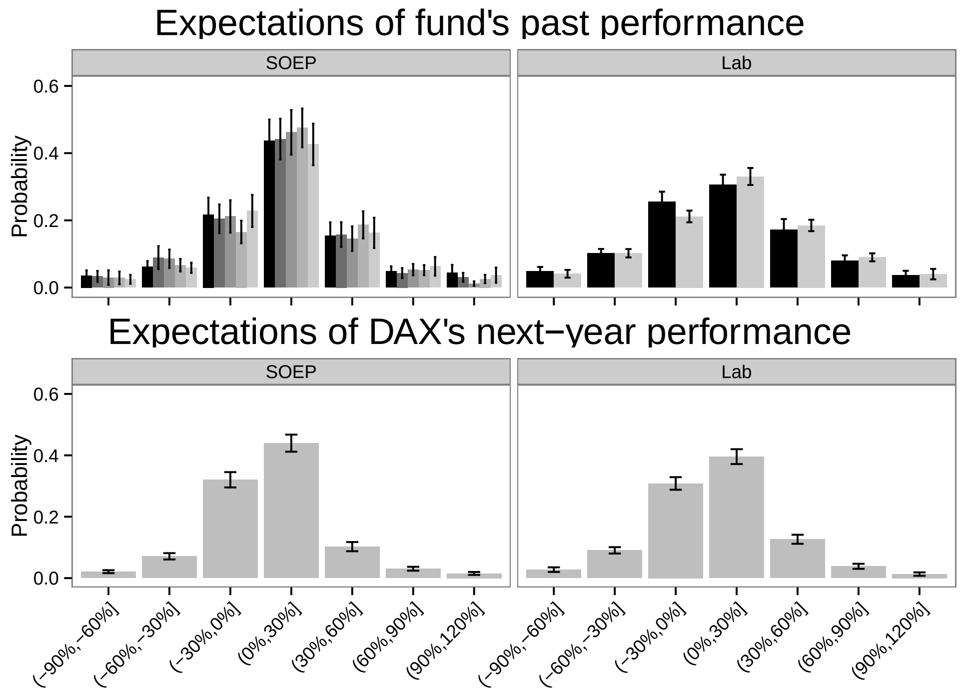
Figure A1: Historical distribution of returns vs. the average distributions in Lab and SOEP

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 grey 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 slighly 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.[32] 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, subtantial 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 corresponsing frequency was 49%. Dominitz and Manski (2011) find similar numbers in the monthly surveys of the Michigan

[32]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 auxilliary 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 substatially (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	Yes	No	No	Yes	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

Figure A3: 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 5, 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 A2 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 heterogenous treatment 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 A2 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 A1 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.

	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***	0.320**	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)	(0.900)	(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 Appendix H.

Table A1: Stock market participation by subgroups

	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 A2: Treatment effect by subgroups

D Histogram Belief Elicitation Screen

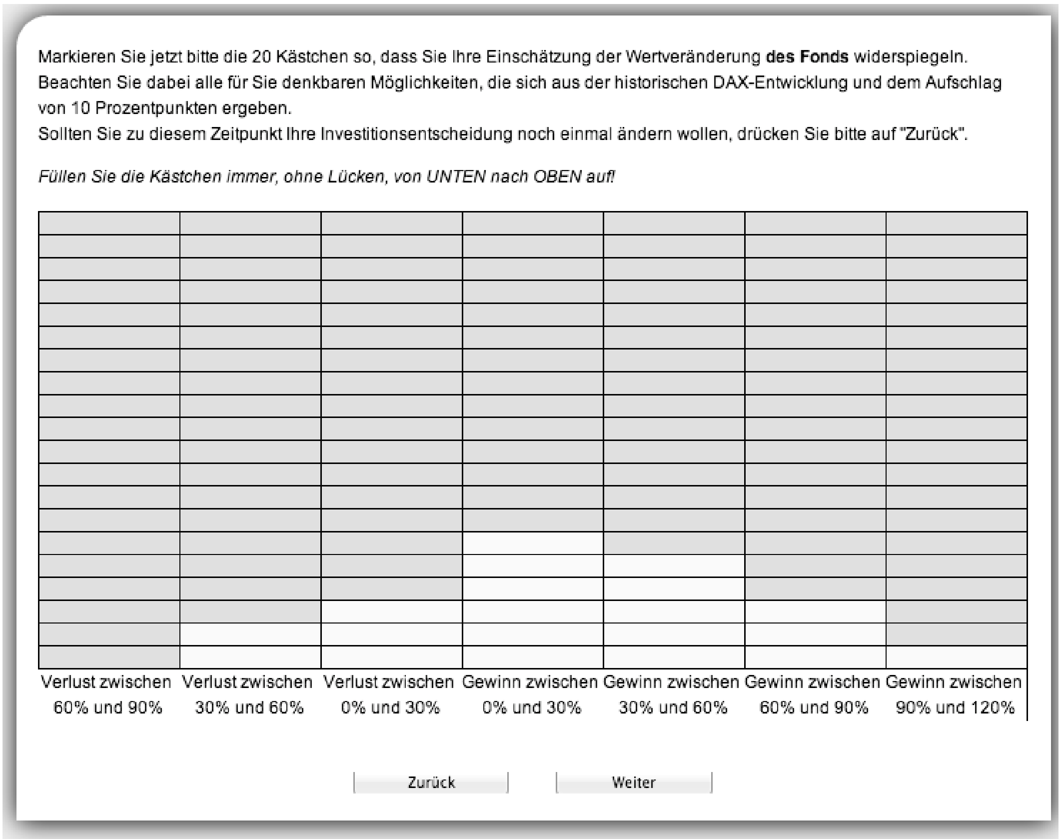


Figure A4: Belief elicitation screen

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	0.320	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	0.320	0.000	1.000
Total Liquid Assets	515	19.000	44.000	0.000	446.000
Stock Market Participation	693	0.180	0.380	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 A3: Descriptive statistics for the 700 heads of household in SOEP sample

F Some Individual Belief Distributions

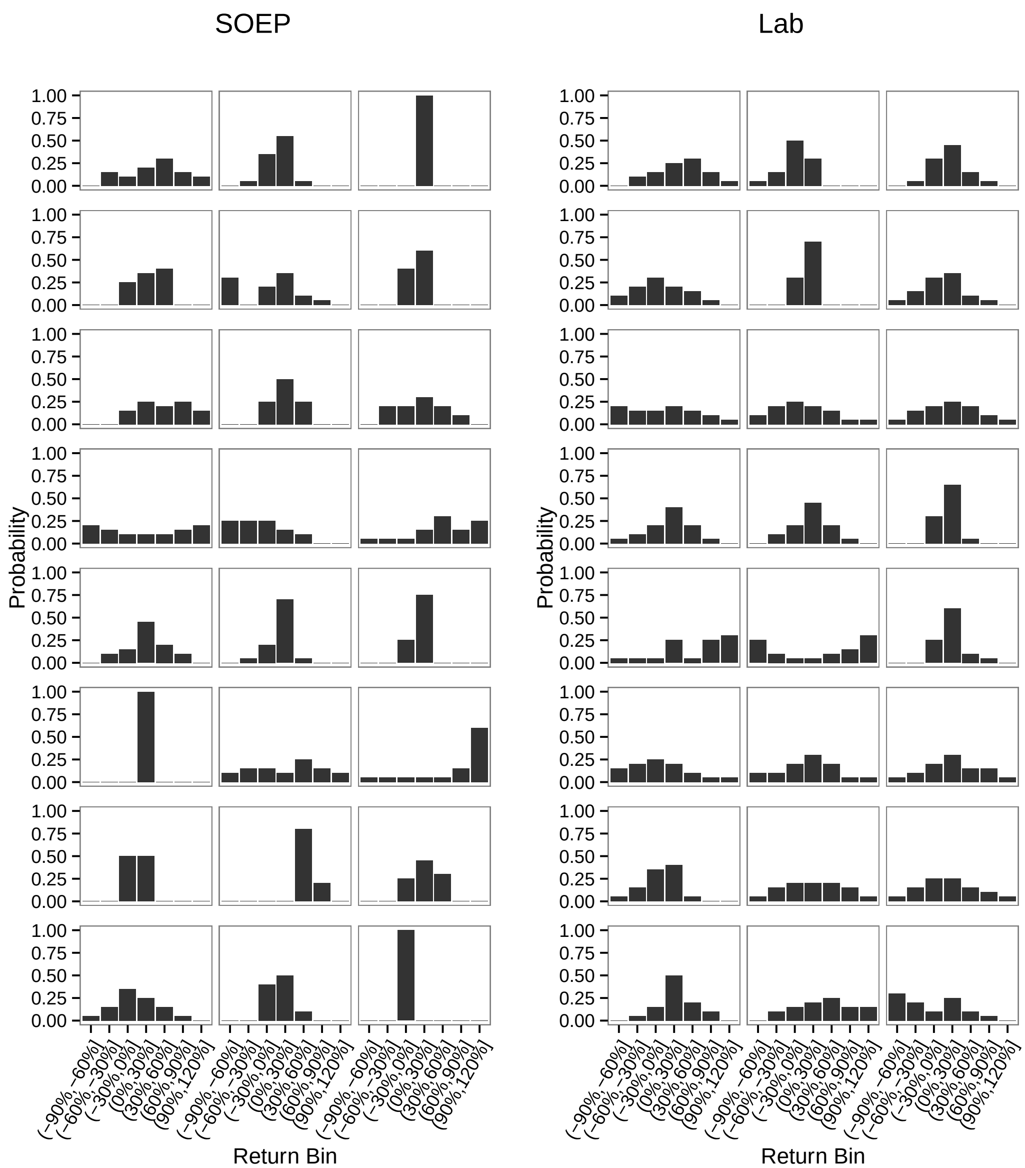


Figure A5: 24 randomly chosen belief distributions from both the SOEP and the lab sample.

G 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.[33]

Figure A6 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

[33]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

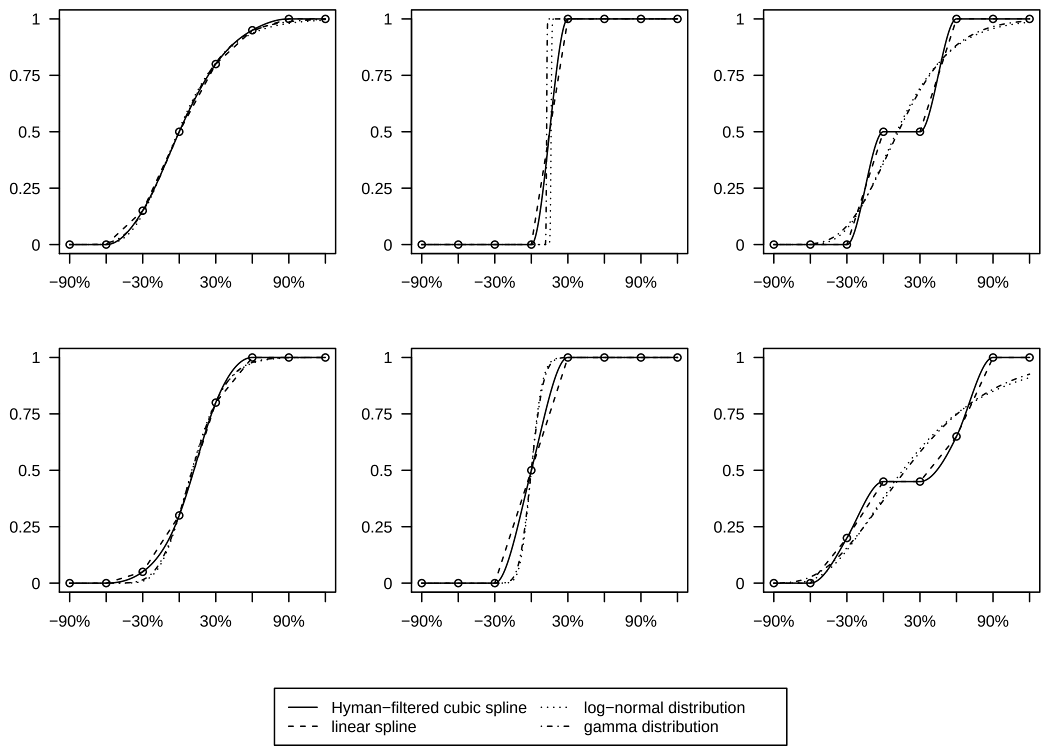


Figure A6: CDFs derived from the belief data using both spline interpolation and parametric distributions fit via least squares

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.

H Variable Description and Coding

The full data set contains 1146 respondents in 700 households. Since asset allocation is commonly seen in the literature as the result of joint optimization of all household members we narrow the sample to the 700 heads of household, which we identify as the respondents who filled out the SOEP household questionnaire. All demographics whose coding is detailed below are the demographics of this household head.

Abitur

Germany has a multi-track educational system in which only students who graduate from high school with an "Abitur" diploma are automatically allowed to enroll at university. In the SOEP respondents are asked directly for the highest secondary school degree they have obtained and our Abitur variable is coded mainly according to the answer to this question. There is one special case, however, that requires special attention. 59 respondents obtained their secondary education outside of Germany and a separate question gives too little information to be able to map the secondary education they obtained into the German educational system precisely. Of these subjects, 11 have university degrees, however; education for which, had it been obtained in Germany, the Abitur would almost always be a prerequisite. Since we are interested in the Abitur as a proxy for higher ability and higher education and foreign respondents with university degrees plausibly posess the same higher ability and higher education we recode these subjects as having Abitur.

Born in East Germany

This indicator variable is 1 if the respondent was born in the German Democratic Republic. It is 0 for respondents born in the Federal Republic of Germany, those born outside of Germany and those born in East Germany after reunification in 1990 (14 cases).

Interest from Wealth

This variable is our main proxy for responents' liquid wealth holdings. Though our survey module included detailed questions about more specific asset classes, item non-response rates for the questions asking for the invested amounts were fairly high. The household questionnaire also included the question "How large, all in all, was your income from interest, dividend payments and capital gains in 2011", with six answer categories.[34] For the econometric analysis we generate a variable that uses information from both questions. We create a new category for subjects who report that their capital income was precisely zero, sort all respondents who gave exact answers into the six categories above and then merged the highest three categories into a single category for capital incomes above €2500 to increase the cell count (counts before the merge were 20 for the €2500 to €5000 category, 5 for the €5000 to €10000 category and 5 for the more than €10000 category). Lastly, we added a category for all subjects who refused to answer both questions.

Financial Literacy

We assess respondents' financial literacy in two different ways. First, we ask people to self-assess their financial literacy with the question:

"**How good, all in all, are you with financial matters?**"[35]

- very good
- good
- a little
- not at all

Second, we ask two questions that explicitly test respondents' financial literacy:

[34] In German: "Wie hoch waren, alles in allem, die Einnahmen aus Zinsen, Dividenden und Gewinnen aus allen Ihren Wertanlagen im Jahr 2011?". Many respondents were either unwilling or unable to provide a precise answer to this question. In a follow-up question they were therefore asked to estimate the amount and choose between 6 categories: below €250, €250 to €1000, €1000 to €2500, €2500 to €5000, €5000 to €10000, more than €10000

[35] In German: "Wie gut kennen Sie sich alles in allem in finanziellen Angelegenheiten aus? Gar nicht, ein bisschen, gut oder sehr gut?"

"**Suppose you have €100 in a savings account. You receive 20% on this amount per year and leave the money in the account for 5 years. How much money will be in the account after these 5 years?**"[36].

- more than €200
- exactly €200
- less than €200
- don't want to answer

"**Which of the following types of investments has the largest fluctuations in returns over time?**"[37].

- savings accounts
- fixed income securities
- stocks
- don't want to answer

Liquid assets

All household members are asked about individual holdings of the following asset types:

1. checking accounts
2. savings accounts
3. call deposit accounts ("Tagesgeld")
4. fixed deposits
5. covered bonds, municipal bonds, bank bonds, corporate bonds or sovereign bonds
6. stock market mutual funds, stocks or reverse convertible bonds ("Aktienanleihen")

[36]In German: "Angenommen, Sie haben 100 €Guthaben auf Ihrem Sparkonto. Dieses Guthaben wird mit 20% pro Jahr verzinst, und Sie lassen es 5 Jahre auf diesem Konto. Wie viel Guthaben weist Ihr Sparkonto nach 5 Jahren auf?"

[37]In German: "Was glauben Sie: Welche der folgenden Anlageformen zeigt im Laufe der Zeit die höchsten Ertragsschwankungen? Sparbücher, festverzinsliche Wertpapiere oder Aktien?"

7. real estate funds
8. bond and money market funds
9. other funds
10. other securities

For each of these types, respondents are first asked whether they own any assets of that type at all and, if the question is answered affirmatively, about the size of the asset holdings. Respondents are instructed to estimate this amount should they be unable to provide an exact figure. We code a household as participating in the stock market if the head of household answers the question about stock market mutual funds, individual stocks and reverse convertible bonds with “yes”.

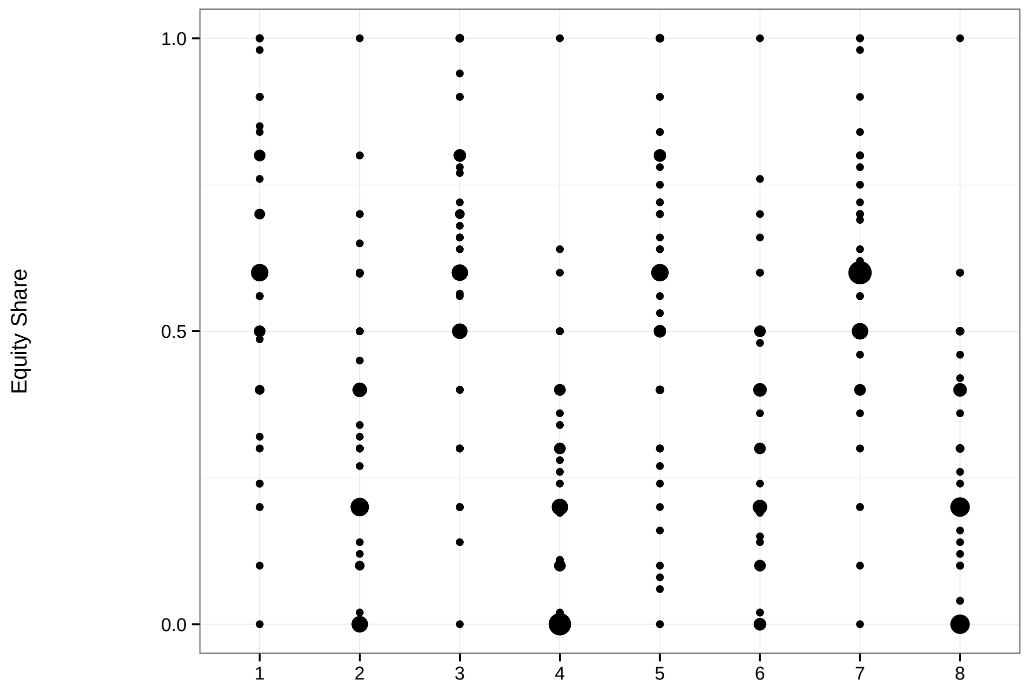
I Robustness Check – Predicting real-world stock-market participation – alternative wealth measures, alternative specifications

	Dependent Variable: Stock Market Participant					
	OLS	OLS	OLS	OLS	OLS	Probit marinal effects
	(1)	(2)	(3)	(4)	(5)	(6)
Equity Share	0.220***	0.240***	0.200***	0.210***	0.140*	0.170***
	(0.072)	(0.068)	(0.064)	(0.066)	(0.076)	(0.056)
Female		−0.043	−0.029	−0.028	−0.028	−0.016
		(0.032)	(0.030)	(0.030)	(0.033)	(0.029)
Born in East Germany		−0.058*	−0.044	−0.032	−0.021	−0.079**
		(0.034)	(0.033)	(0.032)	(0.036)	(0.036)
Age		0.006	0.004	0.002	0.006	0.003
		(0.005)	(0.006)	(0.005)	(0.006)	(0.006)
Age^2		−0.0001	−0.0001	−0.00004	−0.0001	0.000
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.000)
Abitur		0.200***	0.150**	0.140**	0.120*	0.140***
		(0.061)	(0.058)	(0.058)	(0.065)	(0.044)
University Degree		0.049	−0.003	0.013	−0.014	−0.021
		(0.078)	(0.072)	(0.074)	(0.083)	(0.052)
Household Size		0.039**	−0.004	0.003	0.013	0.003
		(0.019)	(0.022)	(0.023)	(0.028)	(0.019)
Risk Tolerance: Low		0.020	0.034	0.033	0.020	0.017
		(0.037)	(0.035)	(0.035)	(0.039)	(0.033)
Risk Tolerance: High		0.008	0.058	0.052	0.058	0.068
		(0.044)	(0.043)	(0.043)	(0.048)	(0.042)
Imputed expectation of DAX		0.001	0.0003	0.0005	−0.0002	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
S.D. of DAX		−0.003***	−0.001	−0.002**	−0.001	−0.002**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gain Probability of DAX		−0.003	0.039	0.035	0.096	0.003
		(0.088)	(0.085)	(0.081)	(0.096)	(0.083)
Number of Children in Household		−0.096***	−0.057*	−0.067**	−0.072**	−0.092***
		(0.030)	(0.030)	(0.031)	(0.035)	(0.031)
Employed		−0.015	−0.024	−0.030	−0.006	−0.015
		(0.036)	(0.037)	(0.037)	(0.042)	(0.039)
Financially Literate		0.140***	0.080***	0.091***	0.078**	0.071**
		(0.032)	(0.031)	(0.032)	(0.036)	(0.030)
Interest: < 250 Euros			0.061*			0.120***
			(0.033)			(0.046)
Interest: 250 - 1.000 Euros			0.270***			0.260***
			(0.057)			(0.047)
Interest: 1.000 - 2.500 Euros			0.430***			0.330***
			(0.086)			(0.058)
Interest: > 2.500 Euros			0.310***			0.270***
			(0.110)			(0.069)
Interest: refused to answer			0.150			0.170*
			(0.100)			(0.090)
Total Liquid Assets (missing=0)				0.011***		
				(0.003)		
Total Liquid $Assets^2$ (missing=0)				−0.0001**		
				(0.00003)		
Total Liquid $Assets^3$ (missing=0)				0.00000		
				(0.00000)		
Total Liquid Assets: missing				0.130***		
				(0.040)		
Household Income (missing=0)			0.023	0.032*		0.020*
			(0.018)	(0.017)		(0.012)
Household Income: missing			0.210**	0.230***		0.180***
			(0.084)	(0.082)		(0.069)
Total Liquid Assets					0.012***	
					(0.003)	
Total Liquid $Assets^2$					−0.0001**	
					(0.00003)	
Total Liquid $Assets^3$					0.00000	
					(0.00000)	
Household Income					0.020	
					(0.019)	
Constant	0.110***	−0.130	−0.130	−0.100	−0.210	
	(0.029)	(0.140)	(0.140)	(0.130)	(0.140)	
N	561	560	560	560	417	560
R^2	0.021	0.150	0.280	0.290	0.310	

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

Income and Liquid assets are in thousands of Euros. Standard errors for OLS regressions are Huber-White heteroskedasticity-robust. Standard errors for probit marginal effects are bootstrapped with 1000 replicates

J Raw Data in Complexity Experiment



Illiquid Endowment	16000.00	16000.00	17550.00	14375.00	14275.00	17275.00	15800.00	15800.00
Bonus on Riskless Asset	5.90	5.90	2.80	9.15	6.05	6.05	3.00	9.00
Bonus on Risky Asset	9.00	2.65	5.90	5.90	9.10	3.10	6.05	6.05

Treatment

Point size is proportional to the number of overlapping observations.

Figure A7: Raw Data in Complexity Experiment