Journal of Econometrics 200 (2017) 378–389

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The precision of subjective data and the explanatory power of economic models

Tilman Drerup [a], Benjamin Enke [b], Hans-Martin von Gaudecker [c,*]

[a] Stanford University, United States
[b] Harvard University, United States
[c] University of Bonn, Department of Economics, Adenauerallee 24-42, 53113 Bonn, Germany

ARTICLE INFO

Article history:
Available online 6 July 2017

JEL classification:
C35
C51
G11

Keywords:
Measurement error
Subjective expectations
Stock market participation

ABSTRACT

Subjective expectations are important primitives in many economic models, yet their direct measurement often yields imprecise and inconsistent data. This has previously been treated as a pure measurement error problem. In contrast, this paper argues that the individual-level precision of such data may reflect the structure of the underlying decision process. We estimate a semiparametric double index model on data specifically collected for this purpose and show that stock market participation decisions exhibit little variation in economic model primitives when individuals provide error-ridden belief statements. In contrast, beliefs and risk preferences predict strong variation in stock market participation for individuals who report precise expectations measures.

1. Introduction

Stock market expectations are among the most important primitives of economic portfolio choice models. With the recent emergence of large-scale datasets including subjective expectations, researchers have begun to incorporate them into empirical models of investor behavior. While the results have been by and large encouraging, working with subjective beliefs data has proved challenging. First, many researchers are troubled by the apparent pervasiveness of measurement error in subjective expectations data. For example, stated beliefs often cluster at focal points (Kleinjans and van Soest, 2014) and many respondents' answers violate even the most basic laws of probability (Manski, 2004; Hurd et al., 2011). Second, the association between subjective beliefs and stockholding decisions tends to be statistically significant, but usually rather small in magnitude (Hurd, 2009; Ameriks et al., 2016).

In this paper, we propose a reconciliation of these two facts. Our point of departure is that different households are likely to employ different thought processes to arrive at their financial decisions. For some people, the canonical economic model of forming a choice rule by combining preferences and beliefs about future states of the world will be a good approximation. Others, however, could take their decisions very differently. For example, almost half of the Dutch population report that they mostly rely on the advice of family, friends, or professionals when it comes to important financial decisions (von Gaudecker, 2015). Likewise, as emphasized by large literatures in behavioral finance and cognitive psychology, households may take financial decisions intuitively (Kahneman, 2011; Binswanger and Salm, 2014) or employ simple rules of thumb (Ameriks and Zeldes, 2004).

The presence of such alternative decision modes can produce the patterns observed in the data. First, individuals who do not base their decisions upon beliefs have little reason to frequently reflect upon the evolution of the stock market. Thus, they will likely maintain only rudimentary, diffuse, and unstable expectations. In consequence, when prodded to state these expectations in surveys, their answers will lack precision: they will be error-ridden, inconsistent, and exhibit large variation across survey instruments. Second, preferences and beliefs will have little explanatory power for portfolio decisions as they do not enter the decision-making process of all individuals. For example, neither preferences nor beliefs will explain variation in the behavior of people who exclusively rely on a rule of thumb to arrive at their financial decisions. In combination, these observations imply our research hypothesis. The responsiveness of financial decisions to variation in subjective expectations and other primitives of economic models should be high for individuals whose stated beliefs exhibit high precision. Beliefs and preferences should induce only very little variation in financial decisions for people with imprecise expectations measures.

To explore the channel of heterogeneous choice rules and motivate our empirical strategy, Section 2.1 presents a simple economic model of stock market participation that clarifies the roles

* Corresponding author.
E-mail addresses: tdrerup@stanford.edu (T. Drerup), enke@fas.harvard.edu (B. Enke), hmgaudecker@uni-bonn.de (H.-M. von Gaudecker).

http://dx.doi.org/10.1016/j.jeconom.2017.06.017
0304-4076/

of expectations, preferences, and transaction costs. In Section 2.2, we discuss in detail why a variety of alternative decision modes imply that individuals have low incentives to frequently reflect upon their beliefs about the future evolution of the stock market.

Section 2.3 lays out our econometric approach. The above arguments suggest that the explanatory power of our model of stock market participation will vary across individuals. To empirically incorporate this particular form of heteroskedasticity, we estimate a Klein and Vella (2009) semiparametric double index model. In this model, the first index contains the primitives of our theoretical model (such as beliefs and preferences), while the second index includes quantitative and qualitative indicators for the precision of measured beliefs. Both indices include further controls and may interact in a fully nonparametric fashion to obtain predicted stockholding probabilities.

Section 3 describes the dataset that we collected specifically for this study. The data contain individual-level information on stock market participation, subjective belief distributions, risk preferences, as well as a variety of quantitative and qualitative proxies for the precision of subjective expectations from a large probability sample of the Dutch population. Section 4 presents the results of our empirical application. We demonstrate that changes in primitives of the economic model induce large variation in stock market participation if expectations measures are precise. If their precision is low, however, the effect of changes in beliefs and preferences on stockholdings is close to zero. We perform a number of variations on this theme and show that the results hold up in several different specifications, such as when we include or exclude proxies for transaction costs into the two indices. We then demonstrate the usefulness of our modeling approach for the analysis of less detailed data by estimating a specification with variables that are commonly available or inexpensive to collect. In particular, we show that restricting ourselves to a simple measure of expectations and purely qualitative proxies for the precision of expectations measures yields a similar, yet less pronounced overall pattern.

Our findings suggest that imprecision in measured beliefs should not necessarily be treated as a standard case of measurement error, which needs to be corrected through, e.g., improved measurement devices or multiple measurements (Wansbeek and Meijer, 2000). While many of the symptoms of diffuse and unstable expectations are observationally equivalent to measurement error, they do not reflect erroneous reporting, but rather the structure of the expectations. Our results hence suggest that individual-level variation in the precision of measured expectations might be informative about economic mechanisms of interest. To bolster this interpretation, we conclude in Section 5 by discussing why our findings are unlikely to be driven by traditional notions of measurement error in subjective beliefs.

2. Motivation and empirical strategy

We develop our empirical strategy in three steps. First, we characterize a household's portfolio choice problem by means of a simple economic model. We then explain in detail why we conjecture that the degree to which this model serves as an adequate description of the decision-making process varies across households and why we expect that variation in the precision of subjective expectations can be exploited to capture this adequacy. In the third step, we present our econometric strategy to implement these ideas.

2.1. *A simple economic model of stock market participation*

Our depiction of households' portfolio choice behavior in an economic model follows Campbell and Viceira (2002). We assume that the household maximizes a power utility function defined over next period's expected financial wealth $E_t\left[W_{t+1}\right]$ by allocating fractions of period-t wealth to one safe and one risky asset. If the household can neither short the risky asset nor leverage his position in it, the optimal risky asset share θ^{opt} solves:

$$\theta^{\text{opt}} = \underset{\theta}{\operatorname{argmax}} \left\{ \frac{E_t\left[W_{t+1}(\theta)^{1-\gamma}\right]}{1-\gamma} \right\} \qquad \text{s.t.} \quad 0 \le \theta \le 1. \tag{1}$$

Risk aversion and a household's beliefs about the returns of the two assets determine the optimal decision. Denote a household's expected return for the safe asset by μ_{t+1}^{safe} and assume that the household's expectations for the risky asset's return can be described by a log-normal distribution with mean μ_{t+1}^{risky} and standard deviation $\sigma_{t+1}^{\text{risky}}$. When returns are log-normally distributed, so is W_{t+1}. For a log-normal variable, it holds that $\log E[X] = E\left[\log X\right] + 1/2\, \text{Var}\left[\log X\right]$. Thus, the maximization problem can be rewritten as follows:

$$\theta^{\text{opt}} = \underset{\theta}{\operatorname{argmax}} \left\{ (1-\gamma) E_t\left[w_{t+1}(\theta)\right] + \frac{1}{2}(1-\gamma)^2 \text{Var}_t\left[w_{t+1}(\theta)\right] \right\} \qquad \text{s.t.} \quad 0 \le \theta \le 1 \tag{2}$$

where lower case letters are logarithms. Using a first-order Taylor series approximation, next period's log wealth can be written as follows:

$$w_{t+1}(\theta) = w_t + (1-\theta)\mu_{t+1}^{\text{safe}} + \theta\mu_{t+1}^{\text{risky}} + \frac{1}{2}\theta(1-\theta)\left(\sigma_{t+1}^{\text{risky}}\right)^2. \tag{3}$$

Substituting this into the expression for θ^{opt} and dividing by $1-\gamma$, we obtain the following expression for the maximand:

$$w_t + \theta\left(\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}\right) + \frac{1}{2}\theta\left(1-\gamma\theta\right)\left(\sigma_{t+1}^{\text{risky}}\right)^2. \tag{4}$$

Solving the first-order condition of this problem for the optimal share θ^{opt} yields

$$\theta^{\text{opt}} = \frac{\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}} + \frac{1}{2}\left(\sigma_{t+1}^{\text{risky}}\right)^2}{\gamma\left(\sigma_{t+1}^{\text{risky}}\right)^2}. \tag{5}$$

At plausible parameter values of γ, the optimal risky asset share will be positive when estimates based on historical return data are used to proxy households' expectations for μ^{safe}, μ^{risky}, and σ^{risky}. However, studies on stock ownership find that a large fraction of the population does not participate in the stock market (e.g., Haliassos and Bertaut, 1995). Arguably the most prominent explanation for why households abstain from participation is the existence of broadly defined transaction costs (Vissing-Jørgensen, 2002). These transaction costs are likely to vary with household characteristics. If participation comes with fixed monetary costs, for example, wealthy households will be more likely to invest in risky assets, since for them the fixed costs are spread over larger investments. If information costs play an important role, transaction costs will be lower for numerate respondents who are quicker to grasp the basic functioning of the stock market. We assume that the variables affecting transaction costs can be modeled by observable household characteristics X^{ta}; denote the resulting transaction costs by $f\left(X^{\text{ta}}\right)$.

We now combine the optimal risky asset share (5), transaction costs, and random influences ε in a simple random utility model of stock market participation:

$$Y \equiv \mathbf{I}\{\theta > 0\} = \begin{cases} 1 & \text{if} \quad \theta^{\text{opt}}\left(\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}, \sigma_{t+1}^{\text{risky}}, \gamma\right) \\ & \qquad -f\left(X^{\text{ta}}\right) > \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

According to (6), the probability of participating in the stock market will depend on the mean and variance of beliefs over the risky asset, the expected risk-free rate, risk aversion, variables proxying transaction costs, and the stochastic properties of ε. If the latter was normally distributed, one could estimate (6) by means of a standard Probit model. Estimators that make minimal distributional assumptions but enable the researcher to recover marginal effects still require ε to either be homoskedastic or have a very particular form of heteroskedasticity (Klein and Vella, 2009). If our conjecture about a varying explanatory power of $\theta^{\text{opt}} - f\left(X^{\text{ta}}\right)$ is correct, this will be reflected in a form of heteroskedasticity that violates these assumptions. In particular, the variance of ε will vary with the precision of beliefs in a form that is unknown a priori.

2.2. Putting the precision of subjective data to productive use

The model combines effortful reasoning about future states of the world with personal risk tolerance to form a choice rule. While such behavior is at the heart of economic thinking, it might not be an adequate description of *all* households' decision processes. Instead, the behavioral finance and cognitive psychology literatures propose that households employ a number of alternative processes to arrive at their decisions. For example, almost half of the Dutch population report that they mostly rely on the advice of family, friends, or professionals when it comes to important financial decisions (von Gaudecker, 2015). Other households may have a tendency to take decisions intuitively (Kahneman, 2011; Binswanger and Salm, 2014) or rely on simple rules of thumb like holding an equity share of 100 minus age (see, e.g., the discussion in Ameriks and Zeldes, 2004).

Many of these alternative decision processes, however, do not require households to frequently reflect about the future evolution of the stock market. As a consequence, we suggest that households who rely on such decision processes are likely to maintain very rudimentary, possibly diffuse, or even unstable expectations. Eliciting such expectations will lead to imprecise, inconsistent, and error-ridden measurements even when the same survey instrument is used at different points in time. Likewise, such respondents should find tasks related to belief elicitation rather difficult and the confidence they express in their estimates should be low.

Indeed, these patterns closely resemble the measurement issues that have been documented in the vast literature on subjective expectations of stock market developments (see the excellent overviews in Manski (2004) and Hurd (2009)). For example, when asked for their expectations about the future of the stock market, respondents frequently violate basic laws of probability or they provide focal point answers such as 50:50 (Bruine de Bruin et al., 2000; Manski, 2004; Hurd, 2009; Bruine de Bruin and Carman, 2012; Kleinjans and van Soest, 2014; Binswanger and Salm, 2014).[1] In addition, non-response tends to be concentrated among subgroups who do not follow the development of the stock market (Hurd, 2009), suggesting that stating beliefs requires significant cognitive effort for people who are not accustomed to reflecting upon the stock market.[2]

Previously, such patterns have frequently been interpreted as cases of measurement error (see, e.g., the discussion in Manski, 2004). However, while often observationally equivalent to measurement error, the semantics of imprecise expectations is very different from the contexts in which measurement error is usually studied. In the case of variables like past income, savings, or consumption, measurement error arises because of, e.g., imperfect recall (Hoderlein and Winter, 2010) or incongruent definitions of precisely defined "true" non-stochastic quantities. In the case of subjective expectations, however, we conjecture that the precision and meaningfulness of expectations measures reflects the structure of beliefs itself, e.g., subjective Knightian uncertainty. In consequence, when attempting to predict household investment behavior, the degree of precision should be informative about the relevance of expectations in the decision process. Specifically, we hypothesize that measures indicative of more precise expectations should be associated with an increase in the explanatory power of expectations for variation in portfolio decisions.

In sum, different pieces of evidence suggest that part of the population holds only imprecise subjective stock market beliefs. We propose that this imprecision contains informational content that will allow us to uncover heterogeneity in choice behavior. In particular, we suggest that the degree of imprecision will allow us to evaluate to which extent households' stock market participation decisions are adequately described by the simple model discussed above.

2.3. Econometric specification

In econometric terms, a consequence of varying precision in expectations measures is that ε in (6) will be heteroskedastic, i.e., its variance will increase as subjective expectations become noisier. Depending on the precise decision-making process, it may also have group-specific means different from zero. For example, the most prevalent advice by family and friends seems to be non-participation in the stock market (von Gaudecker, 2015). For the group of individuals who follow this advice, participation rates will be low even if $\theta^{\text{opt}} - f\left(X^{\text{ta}}\right)$ takes on positive values on average. To capture these consequences, we require an econometric specification where the predictions of the choice model (6) interact with the extent of precision in subjective expectations data in a flexible way.

The double index binary choice model of Klein and Vella (2009) is ideally suited for the structure of our problem. The model obtains an estimate for a binary outcome by nonparametrically combining two linear, partially observable indices; this is precisely what we are interested in. In our case, the indices reflect the economic model and the precision of beliefs. While we observe neither directly, we observe a number of indicators that allow us to approximate each. One can think of this approach as a regime shifting model where one index gradually changes the regime under which the other one operates. While we feel relatively comfortable in approximating both indices in a linear fashion, we do not want to restrict their interaction ex ante because this is the primary mechanism that we hypothesize and want to uncover.

We first aggregate $\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}, \sigma_{t+1}^{\text{risky}}, \gamma$, and X^{ta} into one vector X^{mod}; $X^{\text{mod}}\beta^{\text{mod}}$ approximates our choice model from 2.1.[3] We will refer to $X^{\text{mod}}\beta^{\text{mod}}$ as the economic model index in what follows. A second vector X^{sdp} contains the variables related to the

[1] The interpretation of such patterns in Hudomiet and Willis (2013) is similar to ours. They suggest that answers like 50:50, 0:100, or 100:0 for subjective survival probabilities reflect subjective probability distributions characterized by a high degree of individual-level (Knightian) uncertainty. They recover the distribution of subjective uncertainty from a highly parameterized model, an approach that we view complementary to ours. In our case, one may think of decision rules different from the economic model as giving rise to such uncertainty or vice versa.

[2] Similar patterns of imprecise measurements have been documented for risk preferences. von Gaudecker et al. (2011) and Choi et al. (2014) show that for respondents with high socio-economic status, sequences of lottery decisions are much more consistent with flexible parametric utility functions and the generalized axiom of revealed preferences, respectively. Put differently, risk preference parameters are much more precisely measured for these subgroups.

[3] We also experimented with calculating (5) and including it alongside X^{ta}. This led to numerical difficulties as the covariance matrix of the two indices was near-singular for a wide range of parameter values. We attribute this to the lack of a quantitatively meaningful measure of γ (Rabin, 2000) and to a fat right tail of $\left(\sigma_{t+1}^{\text{risky}}\right)^2$. The latter is likely responsible for the numerical problems; it is also the reason why we use the standard deviation of beliefs instead of the variance.

subjective data's precision. These will be quantitative and qualitative indicators as well as covariates that we would expect to influence the "propensity to use economic reasoning"; we allow the latter to overlap with the transaction cost proxies included in the economic model index. Accordingly, we refer to $X^{\text{sdp}}\beta^{\text{sdp}}$ as the subjective data precision index. The Klein and Vella (2009) estimator models the relationship of both indices and risky asset holdings as follows[4]:

$$P\big(Y=1 \,\big|\, X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big) = h\big(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big). \tag{7}$$

This structure is directly related to (6) in that the subjective data precision index further parameterizes ε, i.e., the random component is systematic to some extent. The function $h(\cdot,\cdot)$ provides a nonparametric link mapping the indices for the economic model and subjective data precision into stock market participation probabilities.

To attain identification (up to location and scale) of the parameters β^{mod} and β^{sdp}, we require that at least one continuous variable per index is excluded from the other index. In each index, we normalize the coefficients on one of these variables. The resulting model satisfies the form in A5 of Klein and Vella (2009) without requiring reparameterization. Under assumptions given in Klein and Vella (2009) – mainly smoothness of $h(\cdot,\cdot)$ and compact support of the covariates – the probability to participate in the stock market can be expressed as a function of the densities conditional on participation:

$$\begin{aligned} &P\big(Y=1 \,\big|\, X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big) \\ &\quad= \frac{f_{Y=1}\big(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big)\cdot P\big(Y=1\big)}{f\big(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big)}, \end{aligned} \tag{8}$$

where $f(\cdot)$ denotes the unconditional density of the bivariate index and $f_{Y=1}(\cdot)$ its density conditional on participation in the stock market. Kernel density estimators for these quantities are obtained under a multi-stage local smoothing procedure to achieve a sufficiently low order of the bias. Denoting the resulting estimator for (8) as $\hat{P}_i\left(\beta^{\text{mod}},\beta^{\text{sdp}}\right)$, we can write the semiparametric maximum likelihood estimator for β^{mod}, β^{sdp} as follows:

$$\begin{aligned} \left(\hat{\beta}^{\text{mod}}_{\text{ml}},\ \hat{\beta}^{\text{sdp}}_{\text{ml}}\right) = &\operatorname*{argmax}_{\beta^{\text{mod}},\ \beta^{\text{sdp}}} \sum_{i=1}^{N} \hat{\tau}_i \Big[Y_i \cdot \log \hat{P}_i\left(\beta^{\text{mod}},\beta^{\text{sdp}}\right) \\ &+ (1-Y_i)\cdot \log\left(1-\hat{P}_i\left(\beta^{\text{mod}},\beta^{\text{sdp}}\right)\right)\Big], \end{aligned} \tag{9}$$

where $\hat{\tau}_i$ denotes a smooth trimming function ensuring that densities do not become too small (Klein and Spady, 1993). Klein and Vella (2009) show that $\left(\hat{\beta}^{\text{mod}}_{\text{ml}},\ \hat{\beta}^{\text{sdp}}_{\text{ml}}\right)$ converges at rate $\sqrt{N}$ to its true value. While the parameter values do not allow for a direct interpretation, various quantities of interest like average partial effects can be computed with little effort.

In sum, when it comes to generating choice behavior, our empirical model allows for a flexible interplay between traditional economic parameters and proxies for their precision. In particular, it will allow an analysis of how marginal changes in model parameters translate into variation in stock market participation, and how this relationship varies across respondents.

3. Data and descriptive statistics

Our data stem from the Dutch LISS study (Longitudinal Internet Studies for the Social Sciences), which regularly administers Internet surveys and experiments to a panel of households comprising a probability sample drawn from the population register kept by Statistics Netherlands.

Implementing our empirical strategy requires data on individual stock market participation, subjective beliefs and risk aversion, proxies for the degree of imprecision in individual responses, and a rich set of sociodemographic covariates. Only the latter are present in the LISS panel by default. In order to obtain measures for the main quantities of interest, we implemented a series of incentivized experiments and survey questions in August and September of 2013. We restricted our experiments to households with financial wealth in excess of 1000 € to focus on respondents with substantial incentives to think about portfolio allocations. To increase turnout, we also included individuals who refused to answer questions about their exact amount of wealth. Within households, we selected the financial decision maker. In total, 2125 individuals completed both survey waves. After dropping observations with missing data, we are left with a final sample of 2072 observations.

3.1. Outcome variable: Stock market participation

LISS routinely collects detailed data on respondents' financial background, including information on asset ownership. To ensure the relevance of elicited beliefs for current portfolio allocations, we asked respondents to update their information on asset holdings in August 2013. For this purpose, we asked them whether they had any type of bank or savings account and/or investments (stocks, bonds, funds, or options). Our outcome variable is a binary index that equals 1 if the respective respondent held any investments, and 0 otherwise. A quarter of the households in our sample holds risky assets (cf. Table 1). This is in the range of values reported for the Netherlands from other datasets and earlier periods (Alessie et al., 2004; van Rooij et al., 2011). In particular, using an administrative dataset from the Netherlands, Knoef et al. (forthcoming) report almost exactly the same rate of stock market participation, providing reassuring evidence for the data quality of our main outcome variable.

3.2. Variables entering the economic model index

Subjective Expectations. In August 2013, we asked respondents to describe their expectations about the one-year return of the Amsterdam Exchange Index (AEX). We employed a variation of the ball allocation procedure developed by Delavande and Rohwedder (2008), which was explicitly designed for usage in Internet experiments. For each individual, the procedure yields an 8-binned histogram for the expectation of the AEX's one-year return. Using the resulting 7 points on the cumulative distribution function, we follow Hurd et al. (2011) and fit a log-normal distribution to obtain individual-level measures for μ^{risky}_{t+1} and $\sigma^{\text{risky}}_{t+1}$. Because our theoretical framework requires expected excess returns, we also asked respondents for a point estimate of the return of a one-year investment into a standard savings account as the most prevalent safe asset. Section A.1.1 of the Internet Appendix contains detailed descriptions of both procedures.

Recent research in the experimental economics literature has shown that financial incentives induce more truthful reporting of beliefs in tasks like ours (see, for example, Palfrey and Wang, 2009; Gächter and Renner, 2010; Wang, 2011). In order to incentivize subjects, we employed the binarized scoring rule of Hossain and Okui (2013) which is incentive-compatible for a wide range of utility functions. As is common practice with large samples like ours, we randomly selected one in ten subjects for actual payment. The maximum earnings per selected subject were 100 € and average earnings equaled 39.66 € conditional on being selected for payment in September 2014.

[4] Klein and Vella (2009) frame their discussion in terms of an estimator for a single-equation binary response model with dummy endogenous variable when no instruments are present. A first application that applies it directly to two indices is given in Maurer (2009).

Table 1
Descriptive statistics.
Source: LISS panel and own calculations. Variables related to the confidence in return estimates, task difficulty, and task obscurity are scaled to range between 0 and 1. Risk aversion is the standardized average of 3 standardized risk aversion proxies. We omit standard deviations of binary variables. The number of observations is 2072.

	Statistic		Index	
	Mean	Std. dev.	Model	Subj. data prec.
Holds risky assets	0.25			
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	−1.18	8.10	×	
Subjective beliefs: μ_{t+1}^{AEX}	2.01	6.19		
Subjective beliefs: $\mu_{t+1}^{\text{sav. acc.}}$	3.18	4.89		
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	6.25	4.01	×	
Risk aversion	0.00	1.00	×	
Absolute difference between belief measures	11.20	13.57		×
Confidence in AEX return estimate	0.46	0.23		×
Confidence in sav. acc. return estimate	0.64	0.24		×
Experimental tasks simple	0.51	0.33		×
Experimental tasks clear	0.69	0.25		×
Financial wealth $\in$ (10,000 €, 30,000 €]	0.27		×	×
Financial wealth $\in$ (30,000 €, ∞)	0.27		×	×
Financial wealth missing	0.18		×	×
Net income $>$ 2500 €	0.46		×	×
Net income missing	0.07		×	×
High education	0.38		×	×
$30 < \text{Age} \leq 50$	0.30		×	×
$50 < \text{Age} \leq 65$	0.34		×	×
$\text{Age} > 65$	0.29		×	×

We relegate a detailed presentation of summary statistics of the belief measures to Section A.1.1 of the Internet Appendix and only discuss some notable features at this point. First, the cross-sectional patterns in our data resemble findings in previous literature (e.g., Manski, 2004; Hurd, 2009; Hurd et al., 2011), e.g., we find that male, richer, and better educated respondents tend to hold more optimistic expectations. Second, though our respondents expect a positive AEX return on average, their expectations are rather pessimistic relative to the AEX's historical return distribution: the mean subjective expectation implied by the distribution is 2.01%, while the AEX returned 7.89% (5.93% inflation-adjusted) on average since 1993. This discrepancy between subjective expectations and historic returns aligns with existing results in the literature, in particular those in Hurd (2009) regarding the AEX. In addition, as Figure A.6 in Section A.1.1 of the Internet Appendix shows, our participants tend to place lower probabilities on extreme returns than what has historically been observed. Finally, and in contrast to the relative pessimism we observe for the AEX, the mean expected return for the savings account, 3.18%, exceeds the rates actually offered at the time of the survey (roughly 1%) by a substantial percentage. In our empirical analyses, we employ the difference between the expected mean return for the AEX and the expected return for the savings account as the empirical analog of the expected excess return.

Risk Preferences. In September 2013, we elicited risk preferences by asking respondents to complete a variant of the "Preference Survey Module", which was developed in Falk et al. (2014) to measure economic preference parameters in large-scale surveys. We further describe it in Section A.1.3 of the Internet Appendix. Respondents first provided a qualitative self-assessment of their willingness to take risks in general and in the financial domain. They then made choices in a series of hypothetical binary lottery tasks. In our main analysis, we employ the average of the three measures' standardized values.

Transaction costs. We include several variables to empirically model the impact of transaction costs on stock market participation decisions. We focus on variables that proxy for variation in transaction costs in the form of either monetary or information costs. If monetary expenses of stock market participation are to some degree fixed – e.g., because banks charge a constant amount for setting up and keeping an investment account – then these costs will be less relevant for wealthy households. We therefore include net household income and financial wealth in the economic index to control for variation in the relevance of monetary transaction costs. If comprehension of the basic functioning of the stock market comes with information costs, then these costs will be lower for more numerate and cognitively able households. Both vary with educational attainment and age (McArdle et al., 2011), which we include as further controls.

3.3. Variables entering the subjective data precision index

Several quantitative and qualitative measures serve to capture the precision of subjective expectations data. We employ variables for (i) the consistency with which participants report their expectations, (ii) the confidence they express in their own beliefs, and (iii) their self-assessment concerning both difficulty and clarity of our survey tasks. On top of such direct proxies, we also include the variables entering transaction costs. Indeed, it is difficult to argue for exclusion restrictions in one direction or another for education, income, financial wealth, or age. To alleviate potential concerns that these proxies for transaction costs drive our results, we will also present a specification in which we exclude all of the corresponding variables from our econometric model.

In September 2013, one month after eliciting the distribution of beliefs, we asked the same set of respondents to provide a point estimate for the one-year ahead return of the AEX. As a quantitative proxy for the precision of households' expectations, we compute the absolute difference between the response to this question and the mean belief from the ball allocation task. We conjecture that large discrepancies between the two estimates indicate that a household entertains only diffuse expectations and is thus unlikely to employ them in actual decision-making.[5]

The first two qualitative proxies relate to the confidence respondents have in their own estimates. Following the elicitation of the point estimates for the expected returns of the AEX and the savings account, we asked respondents to use a slider interface to express their confidence in their own belief on a scale from 0 to

[5] We are not aware of changes in the economic environment between the two surveys that could have induced people to systematically and substantially revise their beliefs. Between August and September 2013, the AEX varied little with closing prices between 362.93 and 382.58.

10, where larger values corresponded to more confidence. We conjecture that respondents maintaining only imprecise expectations will have little faith in their own estimates. For our analysis, we scale answers to both questions to the unit interval.

Both in August and September 2013, we asked subjects to use five-point scales to indicate how clear they found the task descriptions and how simple they considered the belief elicitation itself. We expect that respondents who do not have an elaborate belief distribution find it hard to understand and to complete the tasks. For both questions, we aggregate the responses for August and September and we scale the resulting variables to the unit interval to create two further proxies.

The Internet Appendix A provides a more detailed description and further summary statistics of all proxies. The pairwise correlations between task simplicity, clarity, and the two confidence variables are all positive, whereas all of them are negatively correlated to the absolute difference between the two belief measures. Notably, all of the proxies' correlations to sociodemographic variables conform to our prior expectations. For example, the correlations suggest that highly educated households or households with higher net income entertain more precise expectations, resembling previously-found patterns regarding inconsistent survey responses or item non-response (Manski, 2004; Hurd, 2009).

4. Results

We illustrate the intuition behind our approach using a simple set of OLS regressions. Table 2 contains the results of regressing stock market participation on beliefs and sociodemographic controls. We run separate regressions for households whose two beliefs measures are similar and dissimilar, splitting the sample along the median of the absolute difference in beliefs.

The key differences are found in the first few rows. While all coefficients relating to the parameters of the economic model – mean and standard deviation of subjective beliefs (μ_{t+1}^{AEX} and $\sigma_{t+1}^{\text{AEX}}$) along with our measure of risk aversion – go in the expected direction in both samples, they are substantially larger in absolute value for individuals with more precise beliefs. At the same time, there are no important differences in the transaction cost proxies. These results indicate that the economic model of stock market participation has a higher explanatory power if individuals hold more precise beliefs.

4.1. Main specification

Table 3 presents parameter estimates for the coefficients of the main specification. In the economic model index, we normalize the coefficient on $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$ to 1, thus expressing the remainder of β^{mod} relative to subjective excess return expectations. In the subjective data precision index, we set the coefficient on the absolute difference between the belief measures to -1. Larger values in this index would thus be interpreted as indicative of more precise data. As we will discuss in detail below, the link function $h(\cdot,\cdot)$ is (close to) monotonically increasing in the economic model index as well as in the subjective data precision index. This allows us to infer the direction of partial effects from the coefficient estimates.

The coefficients in both indices are estimated with reasonable precision; their signs and relative magnitudes are plausible given the aforementioned shape of the link function and the scaling of the variables (see Table 1). In particular, all variables with exclusion restrictions have the expected signs and most of them are significant. The economic model index increases in the level of the expected excess returns; it decreases in the standard deviation of returns and in risk aversion. The subjective data precision index increases with all 4 qualitative proxies and, by construction, decreases with the absolute difference between the belief measures.

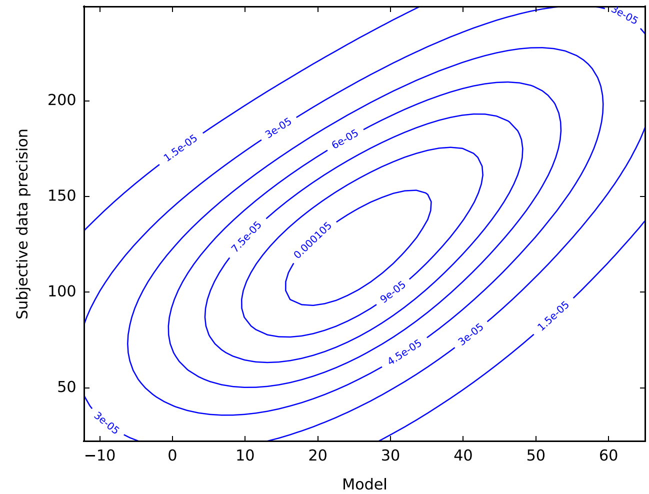

Fig. 1. Joint density of the two indices. LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 2.3 for a detailed description.

Both indices vary significantly with a number of the common covariates. For example, financial wealth is positively related to both indices. This is consistent with wealthy households facing lower transaction costs, while at the same time having stronger incentives to form an opinion about stock market developments. Interestingly, education seems to mostly work through the subjective data precision index, but it has little impact on the economic model index.

For presenting the results of semi- and nonparametric methods, it is particularly important to clarify the support of the data, which in our case refers to the two indices. Fig. 1 shows a contour plot of the joint density of the estimated indices. We limit the area of Fig. 1 and of all subsequent plots to the rectangle spanned by the 5%–95% quantiles of the marginal distributions of both indices. With a correlation coefficient of 0.63, the indices are characterized by a pronounced positive correlation. Note that this correlation does not arise purely mechanically due to the previously noted influence of wealth on both indices – in a model that drops all variables common to both indices (described in the next section), we find the same pattern.

The left panel of Fig. 2 plots the link function $h(\cdot,\cdot)$, i.e., the predicted probability of stock market participation, for varying levels of the economic model and subjective data precision indices. Three features of the plot stand out: First, predicted stock market participation rates vary substantially, ranging from single-digit values to more than 70%. Second, participation rates in general increase monotonically in both the index for the economic model and the subjective data precision index. Third and most importantly, the effects are highly non-linear and interact strongly. In particular, stock market participation is much more responsive to changes in the economic model's ingredients at high levels of the subjective data precision index than at low levels.

To illustrate the last point more clearly, the second panel in Fig. 2 extracts two slices from the first panel. The solid line shows the average response of stock market participation to variation in the model index at the 90%-quantile of the subjective data precision index. There is a pronounced gradient in the middle region, causing predicted risky asset participation to rise from just below 20% to 70%. The dashed line plots the same relation for the 10%-quantile of precision in subjective data. Again, predicted stock market participation varies in the economic model index as expected, but to a much lesser extent. In particular, even for the highest levels of the economic model index, the predicted probability

Table 2
A simple OLS intuition for the double index model.
Source: LISS panel and own calculations. All variables as described in Table 4 in the main text. This table presents regressions of stock market participation decisions on beliefs and sociodemographic controls. The outcome in both columns is a household's decision to participate in the stock market. The left column contains households for which the absolute difference between the two belief measures is smaller than the sample median. The right column contains households where it is larger than the sample median.

	Low abs. diff. in beliefs	High abs. diff. in beliefs
Constant	−0.060	−0.014
	(0.052)	(0.046)
Subjective beliefs: μ_{t+1}^{AEX}	0.012***	0.003*
	(0.003)	(0.002)
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	−0.007*	−0.001
	(0.004)	(0.003)
Risk aversion	−0.078***	−0.043***
	(0.013)	(0.011)
Financial wealth $\in$ (10,000 €, 30,000 €]	0.123***	0.069***
	(0.028)	(0.023)
Financial wealth $\in$ (30,000 €, ∞)	0.444***	0.396***
	(0.033)	(0.037)
Financial wealth missing	0.208***	0.200***
	(0.040)	(0.032)
Net income $>$ 2500 €	0.012	0.061**
	(0.029)	(0.026)
Net income missing	−0.033	−0.000
	(0.046)	(0.047)
High education	0.114***	0.099***
	(0.027)	(0.028)
$30 < \text{Age} \leq 50$	0.043	0.084*
	(0.050)	(0.044)
$50 < \text{Age} \leq 65$	0.126***	0.101**
	(0.049)	(0.043)
Age $>$ 65	0.078	0.053
	(0.051)	(0.041)
Female	0.011	−0.066***
	(0.026)	(0.024)
Married	−0.019	−0.012
	(0.028)	(0.025)
Has children	0.071**	−0.050
	(0.032)	(0.031)
Observations	1041	1031
Adj. (pseudo) R^2 (%)	25.6	23.9

Table 3
Coefficient estimates for the economic model index and the subjective data precision index.
Source: LISS panel and own calculations. The table shows coefficient estimates for the double index binary choice model of Klein and Vella (2009); see Section 2.3 for a detailed description. The dependent variable is a household's stock market participation decision, a binary variable equaling 1 in case the household reports holding any investments, and 0 otherwise. Columns 2 and 3 present estimates of the coefficients and standard errors for the variables contained in the economic model index. Columns 4 and 5 present estimates for the variables contained in the subjective data precision index. In the first index, we normalize the coefficient of the mean excess return to 1, whereas we normalize the coefficient on the absolute difference between the belief measures to -1 in the second index.

	Model		Subjective data precision	
	Estimate	Std. err.	Estimate	Std. err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	–	–	–
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	−0.76	0.29	–	–
Risk aversion	−7.90	1.78	–	–
Absolute difference between belief measures	–	–	−1.00	–
Confidence in AEX return estimate	–	–	59.07	27.55
Confidence in sav. acc. return estimate	–	–	29.25	21.98
Experimental tasks simple	–	–	54.91	19.70
Experimental tasks clear	–	–	15.22	18.20
Financial wealth $\in$ (10,000 €, 30,000 €]	20.17	5.93	19.01	21.22
Financial wealth $\in$ (30,000 €, ∞)	42.75	9.14	91.56	36.81
Financial wealth missing	30.07	7.28	58.27	27.81
Net income $>$ 2500 €	7.32	2.65	−28.49	11.48
Net income missing	−6.37	4.14	4.84	12.86
High education	3.53	2.96	63.62	19.19
$30 < \text{Age} \leq 50$	11.78	5.55	−22.89	16.86
$50 < \text{Age} \leq 65$	7.24	5.53	16.59	15.16
Age $>$ 65	−0.45	5.23	22.81	16.34

of participation does not rise above 30%. This estimated magnitude appears quantitatively large: while the predicted probability of holding stocks only varies by 25 percentage points at the 10th percentile of the subjective data precision index, it varies by more than 50 percentage points at the 90th percentile. The discrepancy in shapes of the two lines highlights the importance of precision

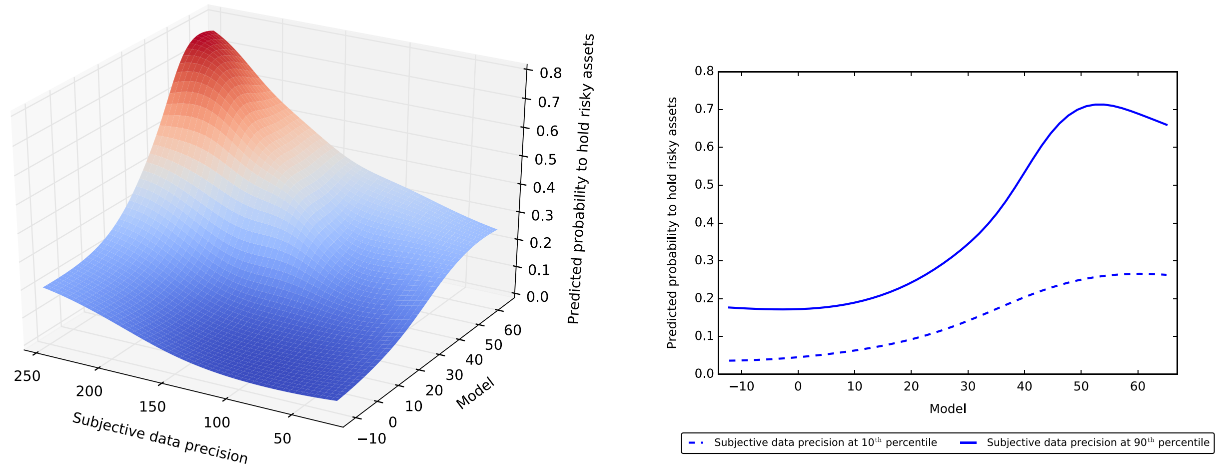

Fig. 2. Predicted probability to hold risky assets. LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision index. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index (43 and 223). Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

in subjective data in understanding the relationship between the primitives of economic models and choices.

We calculate average partial effects to quantify the dependence between individual covariates and stock market participation probabilities. In Table 4, we show how changes in covariates affect participation through either the economic model or subjective data precision index. We also show the combined effect that operates through both indices simultaneously. To calculate average partial effects, we increase continuous variables by one standard deviation. For binary variables, we assign individuals in the left-out category a value of 1.

For the variables solely included in the economic model index, the average partial effects of expected excess return and risk aversion are somewhat larger than the effect of a change in the expected standard deviation of returns. An increase in the expected excess return by one standard deviation is associated with an increase of 3.4 percentage points in the probability to hold investments. Comparable increases in the expected standard deviation and risk aversion reduce the predicted participation rate by 1.4 and 3.8 percentage points, respectively. In the subjective data precision index, a one standard deviation increase in the absolute difference between the two belief measures reduces predicted participation by 1.4 percentage points. Increases in either of the 4 remaining proxies by one standard deviation increase the propensity to participate by between 0.4 and 2 percentage points. If one thinks of the different proxies in terms of a factor structure, varying the underlying factor would likely yield effects of the same order of magnitude as for beliefs or risk aversion.

The effects of financial wealth tend to work through both indices, increasing the propensity to participate in the stock market through the economic model index as well as the subjective data precision index. In contrast, education seems to affect participation mainly through the subjective data precision index.

In sum, this section indicates that respondents' beliefs and risk attitudes are indeed predictive of economic choices. However, the extent to which this is the case varies strongly in the population. Hence, precision in the primitives of the economic model can be used to uncover heterogeneity in its explanatory power.

4.2. Robustness

To illustrate the robustness of our results to alternative specifications of both the economic model and the subjective data precision index, we now present an overview of a number of additional analyses. Section B of the Internet Appendix contains all tables, figures, and additional information.

No transaction cost proxies. Our main specification includes several covariates that proxy transaction costs. Some of them – financial wealth in particular – have strong effects on stock market participation through both the economic model index and the subjective data precision index. To investigate whether the predicted interactions between the economic model and imprecise measures are driven by these sociodemographics only, we estimate one specification without all of the corresponding proxies, i.e., we only include beliefs, risk preferences, and subjective data precision proxies. Except for lower predicted levels of stock market participation at high values of the model index, the overall results on $h(\cdot,\cdot)$ look very similar. Naturally, the partial effects increase.

Mean beliefs only. In this specification, we restrict the model index to consist of expected excess returns only, which gives it an interpretable scale. Section B.2 of the Internet Appendix shows that the gist of our main results is present even in this stripped-down version. The relationship between beliefs and stock market participation is essentially flat at the 10th percentile of the subjective data precision index, while the probability to hold stocks doubles along the beliefs distribution at the 90th percentile of the subjective data precision index. This doubling is concentrated around expected excess returns of zero, whereas the relationship is flat at both extremes of the beliefs distributions. The pattern illustrates the usefulness of our semiparametric approach; typical parametric models such as Logit or Probit would yield the steepest gradient to lie at the right tail of the index' support instead of the center.

Additional covariates. We also check the other extreme and employ a "kitchen-sink"-type approach, including binary variables for gender, having children, and being married in both indices along with the variables from our main specification. It turns out, however, that none of these is significantly associated with either the index of the economic model or the subjective data precision index. In consequence, their inclusion does not affect our results.

Discarding individuals with missing data on financial wealth. In our main specification, we included dummies for financial wealth terciles and for whether information on financial wealth was missing. Since wealth is among the strongest drivers of stock market participation in our model, it is possible that inclusion of respondents with missing information on portfolio value affects our results. To address this concern, we estimate our main specification only with respondents who provided all components of financial

Table 4
Average partial effects.
Source: LISS panel and own calculations. The table presents average partial effects of the Klein and Vella (2009) model; see Section 2.3 for a detailed description. The effects are calculated for a change of 1 standard deviation in continuous variables. For binary variables, we calculate the effect of assigning individuals in the left-out category a value of 1.

	Model	Subj. data prec.	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.034	–	0.034
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	−0.014	–	−0.014
Risk aversion	−0.038	–	−0.038
Absolute difference between belief measures	–	−0.014	−0.014
Confidence in AEX return estimate	–	0.014	0.014
Confidence in sav. acc. return estimate	–	0.008	0.008
Experimental tasks simple	–	0.020	0.020
Experimental tasks clear	–	0.004	0.004
Financial wealth $\in$ (10,000 €, 30,000 €]	0.099	0.017	0.098
Financial wealth $\in$ (30,000 €, ∞)	0.247	0.119	0.373
Financial wealth missing	0.171	0.068	0.219
Net income > 2500 €	0.037	−0.028	0.009
Net income missing	−0.031	0.005	−0.027
High education	0.017	0.080	0.098
$30 < \text{Age} \leq 50$	0.055	−0.025	0.025
$50 < \text{Age} \leq 65$	0.035	0.020	0.054
$\text{Age} > 65$	−0.002	0.027	0.019

wealth. The results are very similar. In particular, the shape of $h(\cdot,\cdot)$ is virtually unchanged. Some of the average partial effects of beliefs and preferences slightly change in magnitude, but all of them qualitatively confirm the main results.

Alternative belief measure. We showed our main results using stated beliefs over the future development of the Amsterdam Exchange Index (AEX). While it is plausible that expectations over a composite index with high media exposure are a good proxy for "the" risky asset in our model, it is still conceivable that our results are biased due to this specific choice. We therefore elicited the same set of belief variables for the future stock return of Philips N.V., one of the largest publicly traded companies of the Netherlands. As one would expect for a single stock with additional idiosyncratic risk, average partial effects relating to the moments of the belief distribution become weaker. The general shape of the link function and the essence of the remaining results, however, is unchanged.

Disaggregated risk aversion measures. By averaging over three distinct variables, we employed a particularly simple aggregation procedure for the risk aversion measure used in our main analysis. When including the three variables separately in the model index, aversion to risk in financial matters emerges as its most important component (Section B.10 of the Internet Appendix). The remainder of our results is not affected.

Interaction between risk aversion and the subjective standard deviation of returns. The main specification contains risk aversion and the standard deviation of the subjective belief distribution as separate variables. To investigate whether increased subjective standard deviation is more important for relatively risk averse subjects, we estimate an additional specification including their interaction in the economic model index. The results in Section B.13 of the Internet Appendix closely resemble those for the main model, and they indicate that the effect of subjective standard deviation does not vary with risk aversion.

Raw returns instead of excess returns. Our theoretical framework suggests employing subjective expected excess returns to predict stock market participation. As discussed in Section 3, our subjects are simultaneously rather pessimistic about the future returns of the market and relatively optimistic about those of a standard savings account. In consequence, a large fraction of our sample expects negative excess returns. While this feature of our data is in line with previous literature, we estimate an additional specification replacing expected excess returns with expected returns to assess the robustness of our results. They are essentially unaffected.

Financial literacy. As mentioned above, a lack of financial literacy may lead subjects to base their participation decision not on expectations about risk and return but on alternative rationales. To assess how our results relate to variation in the respondents' levels of financial literacy, we ran an additional survey in October 2014. In this survey, we asked subjects a set of questions to determine their familiarity with basic financial concepts (Section B.15 of Appendix contains the exact wording). We then used their responses to create binary variables (1 = false answer, 0 = correct answer) and included them in a new specification as additional covariates in both indices. As Section B.15 of Appendix shows, our results remain robust. In addition and in confirmation of our results, most of the average partial effects of the precision proxies are of similar magnitude as in our main specification, suggesting that the precision proxies we employ do not merely pick up a lack of financial literacy.

Alternative ways of calculating the moments of belief distributions. We arrived at our individual-level measures of μ_{t+1}^{AEX} and $\sigma_{t+1}^{\text{AEX}}$ by fitting log-normal distributions to respondents' stated cumulative distribution functions. We obtain very similar results when we estimate the moments assuming uniformly distributed expectations within bins (Section B.11 of the Internet Appendix) or when we follow Bellemare et al. (2012) in approximating each respondent's distribution using a spline interpolation method (Section B.12).

Alternative ways of calculating the absolute difference between belief measures. We constructed a quantitative proxy for imprecise measures as the absolute difference between the point estimate and the mean of the subjective belief distribution. Some subjects, however, may have had the mode or median in mind when providing a point estimate (Delavande and Rohwedder, 2011). Sections B.3 and B.4 of the Internet Appendix show that we obtain quantitatively and qualitatively very similar results when we define the absolute difference based on the median or mode of the belief distribution. To give respondents the benefit of the doubt, Section B.5 estimates one specification where we pick the moment (mean, median, mode) of the belief distribution that minimizes the absolute difference to the point estimate. Again, our findings are not affected.

4.3. Specification with less customized data

Our analyses employ very detailed data on respondents' stock market expectations based on an incentivized Online Experiment. Our proxies for the precision of expectations include a quantitative variable derived from repeated belief measurements and several qualitative indicators. In many surveys, asking for information this detailed is either impossible or impractical. We now evaluate the applicability of our empirical approach to situations with less customized data.

In the model index, we replace the mean of the log-normal belief distribution derived from the ball allocation task by individuals' point estimates. We drop the standard deviation of beliefs and use aversion towards risks in general instead of our composite variable (see Section A.1 of the Internet Appendix for a detailed description of all measures). In the subjective data precision index, we only keep the answers to the qualitative questions which asked respondents about the difficulty and clarity of our survey. We retain all sociodemographic covariates. We then re-run our main analyses using this limited set of variables.

Fig. 3 illustrates that the main results for this model are broadly similar to those of our main specification.[6] As the left panel indicates, the predicted probability of holding risky assets strongly varies with both model indices. Importantly, we find strong variation in the gradient of the economic model even with these much coarser data: While the probability of investing in the stock market is sensitive to changes in the economic model index at high values of the data precision index, the relationship is essentially flat for low levels. The average partial effects in Table 5 again suggest that beliefs and willingness to take risks positively affect stock market participation. The same holds for the precision proxies. All magnitudes are roughly similar to our main specification.

These results entail two consequences: On the one hand, they suggest that imprecise measures will also interfere with our understanding of stock market participation decisions when working with simple measures of beliefs and risk preferences. On the other hand, they suggest that our empirical approach to making productive use of imprecise measures of this kind does not seem to rely on very detailed data to work.

5. Discussion and conclusions

Attempts to measure subjective stock market expectations have dramatically increased over the last two decades. By and large, the results have been encouraging, but obvious signs of poor data quality remain for large fractions of the population regardless of particular survey devices (Manski, 2004; Hurd, 2009; Kleinjans and van Soest, 2014). When these measures have been employed to predict portfolio choice behavior (e.g., Hurd and Rohwedder, 2011; Hurd et al., 2011; Kézdi and Willis, 2011; Hudomiet et al., 2011; Huck et al., 2014), significant correlations in the expected direction have emerged. Nevertheless, it seems fair to say that these are not of the magnitude economists might have hoped for. For example, the abstract of Ameriks et al. (2016) notes that "estimated risk tolerance, expected return, and perceived risk have economically and statistically significant explanatory power for the distribution of stock shares. Relative to each other, the magnitudes are in proportion with the predictions of benchmark theories, but they are all substantially attenuated." In this paper, we have explored a mechanism that can explain both facts. We have argued that differences in the "propensity to use economic reasoning" may drive heterogeneity in the precision of subjective expectations data and explain why the explanatory power of portfolio choice models has been moderate on average.

While the idea of heterogeneous decision rules is certainly not new (e.g., Ameriks and Zeldes, 2004; Kahneman, 2011; Binswanger and Salm, 2014 among many others), we are the first to suggest that the degree of precision in subjective expectations data can be used to uncover such heterogeneity. To explore this link empirically, we have used a semiparametric double index model due to Klein and Vella (2009) on a dataset specifically collected for this purpose. Our results show that stock market participation reacts strongly to the primitives of an economic model (preferences, beliefs, and transaction costs) when subjective data are measured with high precision. When measurement precision is low, there is hardly any reaction at all. This pattern obtains in a wide variety of specification choices, including a setting where we restrict ourselves to variables that are available in many datasets.

A key implication of our findings is that "low quality" of subjective beliefs data should not be treated as a standard measurement error problem, because the strong variation in the precision or meaningfulness of expectations measures actually reflects behaviorally relevant heterogeneity in choice behavior, rather than erroneous reporting. Three pieces of evidence lend further support to our interpretation of the results as reflecting heterogeneous decision modes rather than attenuation bias resulting from standard measurement error. First, if we were dealing with standard versions of measurement error in beliefs (e.g., due to carelessness of some respondents in filling out the survey), taking averages of multiple measurements with uncorrelated idiosyncratic variation should increase the predictive power of expectations. A simple exercise shows that such a pattern does not obtain in our data. We run OLS regressions of stock market participation on convex combinations of our two belief measures (the results are unchanged if we add controls). In Section D of the Internet Appendix, we show that the maximum R^2 is reached close to the point where all the weight is on the mean from the ball allocation task. Hence, adding the second measure hardly helps at all. Second, in all our specifications the likelihood to participate in the stock market was lower for households entertaining imprecise expectations. This suggests that the patterns we found do not merely reflect attenuation bias due to respondents' carelessness or differential effort when responding to the belief questions. If some subjects gave random answers which were uncorrelated with portfolio allocations, participation rates should be the same on average. Third, Armantier et al. (2015) show related patterns for subjective inflation expectations in an experimental setting – financially literate individuals react much more strongly to their expectations than others. Similarly, in an experimental portfolio choice problem, Huck et al. (2014) show that the investment behavior of less sophisticated households is less responsive to exogenous changes in incentives.

Our method is applicable to a wide range of settings where subjective data are used, as long as the dataset contains some individual-level information on the precision or meaningfulness of the respective variables. For example, we noted above that the precision of individual-level risk preference parameters obtained from experiments via revealed-preference paradigms varies tremendously in heterogeneous populations (von Gaudecker et al., 2011; Choi et al., 2014). We have shown how the individual-level precision in data on structural parameters can be used when these parameters are employed to explain economically interesting outcomes. Doing so should help dampen the hostility of economists to subjective data (Manski, 2004) that has arisen largely because of perceived data quality. We have turned this argument around and shown that once there is direct information on data precision at the individual level, it can be used to learn about the economic mechanism of interest.

Acknowledgments

We would like to thank the team of CentERdata, especially Miquelle Marchand, for their help in the data collection process, Philipp Kloke and Lukas Wendlik for able research assistance, Bas van Heiningen for help in recording the introductory screencast, as well as Jürgen Maurer for sharing and explaining the code used in Maurer et al. (2010) that forms the basis of ours. The comments

[6] Section C of the Internet Appendix provides the full set of figures and tables for this model with reduced data requirements.

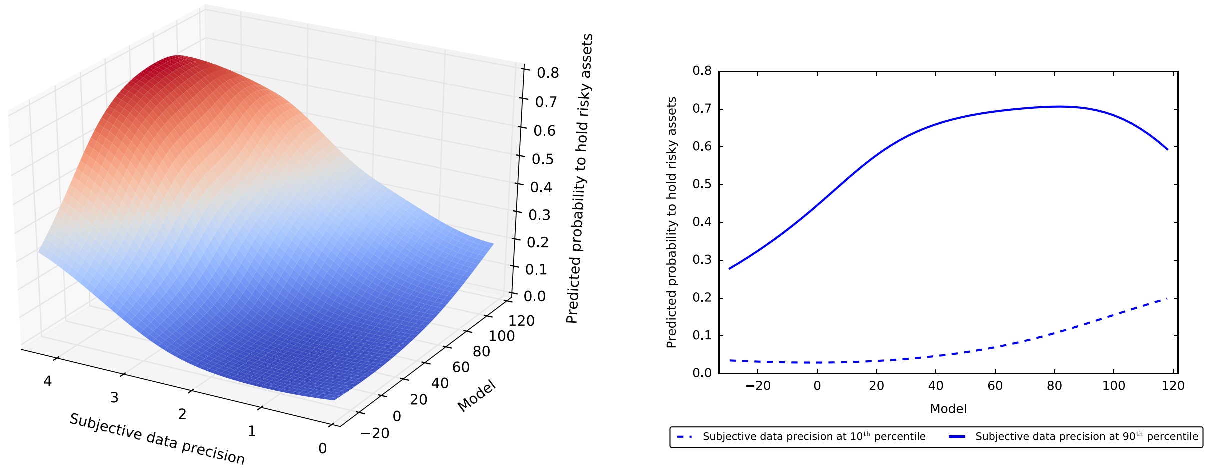

Fig. 3. Predicted probability to hold risky assets, specification with less customized data. LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. These estimations are based on a limited set of variables. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Table 5
Average partial effects, specification with less customized data.
Source: LISS panel and own calculations. The table presents average partial effects of the Klein and Vella (2009) model with a limited number of variables. The effects are calculated for a change of 1 standard deviation in continuous variables. For binary variables, we calculate the effect of assigning individuals in the left-out category a value of 1.

	Model	Subj. data prec.	Combined
Subjective beliefs (direct question): Log expected excess return	0.031	–	0.031
Aversion to risks in general	−0.029	–	−0.029
Experimental tasks simple	–	0.036	0.036
Experimental tasks clear	–	0.010	0.010
Financial wealth ∈ (10,000 €, 30,000 €]	0.077	0.036	0.101
Financial wealth ∈ (30,000 €, ∞)	0.056	0.346	0.396
Financial wealth missing	0.091	0.114	0.202
Net income > 2500 €	0.026	−0.008	0.018
Net income missing	−0.096	0.055	−0.050
High education	−0.001	0.116	0.115
$30 < \text{Age} \leq 50$	0.095	−0.085	0.013
$50 < \text{Age} \leq 65$	0.053	0.014	0.070
Age > 65	−0.035	0.063	0.024

of several referees substantially improved the paper. Seminar participants at the Max Planck Institute for Research on Collective Goods, at Queen's University, at the Mannheim meeting of Young German Microeconometricians, at Tilburg University, and at the Strasbourg conference in honor of Françis Laisney also provided helpful comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2017.06.017.

References

Alessie, Rob, Hochgürtel, Stefan, van Soest, Arthur, 2004. Ownership of stocks and mutual funds: a panel data analysis. Rev. Econ. Stat. 86 (3), 783–796.
Ameriks, John, Keźdi, Gábor, Lee, Minjoon, Shapiro, Matthew D., 2016. Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares. Mimeo, University of Michigan.
Ameriks, John, Zeldes, Stephen P., 2004. How do household portfolio shares vary with age? Mimeo, Columbia Graduate School of Business.
Armantier, Olivier, Bruine de Bruin, Wändi, Topa, Giorgio, van der Klaauw, Wilbert, Zafar, Basit, 2015. Inflation expectations and behavior: do survey respondents act on their beliefs? Internat. Econom. Rev. 56 (2), 505–536.
Bellemare, Charles, Bissonnette, Luc, Kröger, Sabine, 2012. Flexible approximation of subjective expectations using probability questions. J. Bus. Econom. Statist. 30 (1), 125–131.
Binswanger, Johannes, Salm, Martin, 2017. Does Everyone Use Probabilities? The Role of Cognitive Skills. Eur. Econ. Rev. 98, 73–85.
Bruine de Bruin, Wändi, Carman, Katherine G., 2012. Measuring risk perceptions: What does the excessive use of 50% mean? Med. Decis. Making 32 (2), 232–236.
Bruine de Bruin, Wändi, Fischhoff, Baruch, Millstein, Susan G., Halpern-Felsher, Bonnie L., 2000. Verbal and numerical expressions of probability: "it's a fifty–fifty chance". Organ. Behav. Hum. Decis. Process. 81 (1), 115–131.
Campbell, John Y., Viceira, Luis M., 2002. Strategic asset allocation. In: Clarendon Lectures in Economics, Oxford University Press.
Choi, Syngjoo, Kariv, Shachar, Müller, Wieland, Silverman, Dan, 2014. Who is (more) rational? Amer. Econ. Rev. 104 (6), 1518–1550.
Delavande, Adeline, Rohwedder, Susann, 2008. Eliciting subjective probabilities in internet surveys. Publ. Opin. Quart. 72 (5), 866–891.
Delavande, Adeline, Rohwedder, Susann, 2011. Individuals' uncertainty about future social security benefits and portfolio choice. J. Appl. Econometrics 26 (3), 498–519.
Falk, Armin, Becker, Anke, Dohmen, Thomas, Huffman, David, Sunde, Uwe, 2014. An Experimentally Validated Preference Survey Module. Mimeo, Universität Bonn.
Gächter, Simon, Renner, Elke, 2010. The effects of (incentivized) belief elicitation in public goods experiments. Exp. Econ. 13, 364–377.
Haliassos, Michael, Bertaut, Carol C., 1995. Why do so few hold stocks? Econ. J. 105 (432), 1110–1129.
Hoderlein, Stefan, Winter, Joachim, 2010. Structural measurement errors in nonseparable models. J. Econometrics 157 (2), 432–440.
Hossain, Tanjim, Okui, Ryo, 2013. The binarized scoring rule. Rev. Econom. Stud. 80 (3), 984–1001.
Huck, Steffen, Schmidt, Tobias, Weizsäcker, Georg, 2014. The Standard Portfolio Choice Problem in Germany. Mimeo, Humboldt Universität Berlin, Available at http://dx.doi.org/10.2139/ssrn.2432045.
Hudomiet, Péter, Kézdi, Gábor, Willis, Robert J., 2011. Stock market crash and expectations of american Households. J. Appl. Econometrics 26 (3), 393–415.
Hudomiet, Péter, Willis, Robert J., 2013. Estimating second order probability beliefs from subjective survival data. Decis. Anal. 10 (2), 152–170. http://dx.doi.org/10.1287/deca.2013.0266.

Hurd, Michael D., 2009. Subjective probabilities in household surveys. Annu. Rev. Econ. 1 (1), 543–562.
Hurd, Michael D., Rohwedder, Susann, 2011. Stock Price Expectations and Stock Trading, RAND Labor and Population Working Paper, Vol. 938.
Hurd, Michael D., van Rooij, Maarten C.J., Winter, Joachim, 2011. Stock market expectations of dutch Households. J. Appl. Econometrics 26 (3), 416–436.
Kahneman, Daniel, 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York, NY.
Kézdi, Gábor, Willis, Robert J., 2011. Household Stock Market Beliefs and Learning, NBER Working Paper 17614.
Klein, Roger W., Spady, Richard H., 1993. An efficient semiparametric estimator for binary response models. Econometrica 61 (2), 387–421.
Klein, Roger W., Vella, Francis, 2009. A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. J. Appl. Econometrics 24 (5), 735–762.
Kleinjans, Kristin J., van Soest, Arthur, 2014. Rounding, focal point answers and nonresponse to subjective probability questions. J. Appl. Econometrics 29 (4), 567–585.
Knoef, Marike, Been, Jim, Alessie, Rob, Caminada, Koen, Goudswaard, Kees, Kalwij, Adriaan, 2016. Measuring retirement savings adequacy: developing a multi-pillar approach in the netherlands. J. Pension Econ. Finance 15 (1), 55–89.
Manski, Charles F., 2004. Measuring expectations. Econometrica 72 (5), 1329–1376.
Maurer, Jürgen, 2009. Who has a clue to preventing the flu? Unravelling supply and demand effects on the take-up of influenza vaccinations. J. Health Econ. 28 (3), 704–717.
Maurer, Jürgen, Klein, Roger W., Vella, Francis, 2010. Subjective health assessments and active labor market participation of older men: evidence from a semi-parametric binary choice model with nonadditive correlated individual-specific effects. Rev. Econ. Stat. 93 (3), 764–774.
McArdle, John J., Smith, James P., Willis, Robert J., 2011. Cognition and economic outcomes in the health and retirement survey. In: Wise, David A. (Ed.), Explorations in the Economics of Aging. University of Chicago Press, pp. 209–233.
Palfrey, Thomas R., Wang, Stephanie W., 2009. On eliciting beliefs in strategic games. J. Econ. Behav. Organ. 71 (2), 98–109.
Rabin, Matthew, 2000. Risk aversion and expected utility theory: a calibration theorem. Econometrica 68 (5), 1281–1292.
van Rooij, Maarten C.J., Lusardi, Annamaria, Alessie, Rob, 2011. Financial literacy and stock market participation. J. Financ. Econ. 101 (2), 449–472.
von Gaudecker, Hans-Martin, 2015. How does household portfolio diversification vary with financial sophistication and financial advice? J. Finance 70 (2), 489–507.
von Gaudecker, Hans-Martin, van Soest, Arthur, Wengström, Erik, 2011. Heterogeneity in risky choice behaviour in a broad population. Amer. Econ. Rev. 101 (2), 664–694.
Vissing-Jørgensen, Annette, 2002. Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures, NBER Working Paper 8884.
Wang, Stephanie W., 2011. Incentive effects: the case of belief elicitation from individuals in groups. Econom. Lett. 111 (1), 30–33.
Wansbeek, Tom J., Meijer, Erik, 2000. Measurement Error and Latent Variables in Econometrics. Elsevier North Holland, Amsterdam.