Heterogeneity in households' stock market beliefs*

Levels, dynamics, and epistemic uncertainty

Hans-Martin von Gaudecker and Axel Wogrolly

January 2020

Abstract

We analyse a long panel of households' stock market beliefs to gain insights into the nature of the levels, dynamics, and informativeness of these expectations. In a first step, we classify respondents into one of five groups based on their beliefs data alone. In a second step, we estimate models of expectations at the group level so that belief levels, volatility, and response to information can vary freely across groups. At opposite extremes in terms of optimism we identify pessimists who expect substantially negative returns and financially sophisticated individuals whose expectations are close to the historical average. Two groups expect average returns around zero and differ only in how they respond to information: Extrapolators who become more optimistic following positive information and mean-reverters for whom the opposite is the case. The final group is characterised by its members being unable or unwilling to quantify their beliefs about future returns.

Keywords: Stock market expectations, household finance, heterogeneity, clustering methods

JEL codes: D83, D84, D14, C38

*Hans-Martin von Gaudecker: Universität Bonn, 53012 Bonn, Germany. `hmgaudecker[at]uni-bonn[dot]de`, tel. +49 228 73 9357; Axel Wogrolly: Universität Bonn, 53012 Bonn, Germany. `axelwogrolly[at]uni-bonn[dot]de`. We would like to thank Stéphane Bonhomme, Andreas Dzemski, Charles Manski, Simas Kucinskas, and Joachim Winter as well as participants of the first CRC-TR 224 conference, the Ifo/LMU/CRC-TR 190 Workshop on Subjective Expectations and Probabilities in Economics, and the 2018 Econometric Society European Meetings, for their comments. We thank Matthias Vollbracht from Media Tenor International for giving us access to data and Andreas Dzemski for sharing and explaining the code calculating the test statistic from Dzemski and Okui (2018). Both authors are grateful for financial support by the German Research Foundation (DFG) through CRC-TR 224 (Project C01).

1 Introduction

Understanding households' stock market expectations is critical for models of life-cycle behaviour, portfolio choice, and asset pricing. A number of key facts have been established for the cross-section of subjective beliefs about equity returns (Manski, 2004; and Hurd, 2009, provide excellent overviews, we pay detailed credit below). Beliefs differ widely across individuals. On average, they tend to be pessimistic relative to historical returns. Stated beliefs exhibit focal point responses; when it comes to probabilities, 50:50 is a particularly common answer. Stated expectations of a sizeable fraction of individuals are not consistent with the laws of probability. Optimism and consistency of beliefs are positively related to socio-economic variables in general and measures of financial sophistication in particular.

More recently, additional attention has been paid to the process of belief formation as a potential source of this heterogeneity. Taking a long-term perspective, Malmendier and Nagel (2011) show that individuals who experienced larger stock returns over the course of their lives tend to expect larger future returns. Greenwood and Shleifer (2014) find that on average, beliefs extrapolate recent stock market performance into the future. Adam, Marcet, and Beutel (2017) test the rational expectations hypothesis using subjective expectations data and reject it. Barberis et al. (2015) and Adam, Marcet, and Nicolini (2016) develop asset pricing models that feature investors with non-standard belief formation processes, showing that this matters for aggregate outcomes.

Starting from these sets of observations, this paper estimates processes for the formation of households' stock market beliefs, taking into account heterogeneity in levels, volatility, response to information, and epistemic uncertainty. We make use of an unusually long panel of probabilistic belief statements in the RAND American Life Panel, which was commissioned by and first described in Hurd and Rohwedder (2011). We start by verifying in our data the key facts in the cross section and on average belief formation, expanding upon them in several directions. Most importantly, we add the tone of recent media reports on the economy in U.S. television as an additional source of information. We do so because respondents overwhelmingly cite the state of the economy as a driver of their return expectations while at the same time, many claim to not follow the stock market and report incorrect values for realised returns, making it unlikely that the behaviour of stock prices is their prime source of information.

Our analysis of belief heterogeneity focuses on four dimensions: Levels, volatility, response to recent stock market returns and economic news, and epistemic uncertainty. The average time series dimension of our data is 26, which is too short for estimation at the individual level. In order to allow for heterogeneity along the four dimensions, we employ the discretisation approach proposed in Bonhomme, Lamadon, and Manresa (2017). In a first step, we use the k-means clustering algorithm to assign individuals to groups based on the dependent variable. We use a number of individual-level moments relating to levels and volatility, its covariances with recent stock market returns and economic news, and measures of consistency and self-stated information content. These variables thus capture the four dimensions of interest. The procedure yields groups that are similar in spirit to the types studied in Dominitz and Manski (2011) and in Heiss et al. (2019), whose beliefs also differ in their levels, volatility and response to recent stock market returns.

We focus on five groups in our main specification. Using less groups mixes individuals with very different economic behaviours; adding more leads to relatively little additional insights at the expense of making the results harder to summarise. We show that our groups are stable across specifications; varying important features of the sample or of the classifying procedure changes little. Results of the diagnostic tests for group membership by Dzemski and Okui (2018) further corroborate our choice of groups and modelling strategy. All groups are reasonably large with sizes ranging between 13% and 26% of the sample.

In a second step, we estimate models relating respondents' beliefs about future stock prices to past returns of the Dow Jones and the tonality of economic news, allowing parameters to fully vary across groups. We find that one group consists of individuals whose expectations are close to the historical performance of the stock market and who respond slightly positively to recent returns and news about the economy. They have very low rates of inconsistencies. This behaviour is the closest we get to rational expectations; we thus label them "sophisticates". Correlating group membership with other observables, they stand out for having better knowledge of financial markets and the stock market in particular. At the other extreme of average expectations, we estimate one group with substantially negative return expectations and little reaction to returns or news. We label them "pessimists"; they have average values for inconsistencies in the belief elicitation procedure. The latter also is true for two more groups who both have return expectations around zero. Of all groups, these two react the strongest to both returns and news, but in completely different ways. One expects recent trends to continue ("extrapolators"); the other expects them to revert again ("mean reverters"). The last group stands out from the rest in that its members frequently give 50:50 answers when asked about probability judgements and state that these are their way of expressing epistemic uncertainty in a follow-up question; their belief measures often violate the laws of probability calculus. We label this group "ignorants"; correlations with other characteristics reveal that its members indeed do not pay much attention to the stock market.

We show that the groups we identify have very different levels of stockholding and trading behaviour. The level of heterogeneity in trading profiles over our sample period arises because of our classification into groups based on (time-series) features of the dependent variable. Our findings are robust to a number of choices regarding the treatment of the data and to parameters of the classification procedure. Our approach of first grouping indivduals based on the dependent variable and estimating group-level models achieves much higher goodness of fit than using observables alone in a classical regression analysis. This is consistent with recent evidence from a mixed survey-administrative dataset in Giglio et al. (2019), who document persistent heterogeneity in the levels of beliefs that is difficult to explain with observable characteristics.

In a final step, we use the method of Coibion and Gorodnichenko (2015) to test whether the expectations of any of our groups could be characterised as rational in the sense that their forecast errors are unpredictable. We find that this is not the case; all overreact to current information. This is in line with Bordalo et al. (2018) who find evidence of overreaction for a range of macroeconomic variables, and unsurprising in light of how difficult it is to predict stock return better than the historical average does.

The rest of the paper is organised as follows. Section 2 describes our data, connects it to prior literature, establishes the key stylised facts for our data, and outlines our empirical strategy. In section 3, we present the results, including the descriptions of several robustness analyses, the details of which are relegated to the Online Appendix. Section 4 concludes.

2 Data, stylised facts, and empirical strategy

We analyse data from the RAND American Life Panel (ALP, see `https://alpdata.rand.org`) that were collected between 2008 and 2016. The ALP is a panel representative of the U.S. population whose members are regularly interviewed over the Internet. Households lacking internet access upon recruitment were provided laptops to limit selection bias. In addition to providing a large set of background characteristics from regular surveys, the ALP serves as a laboratory for researchers who are able to collect data at low costs. Hurd and Rohwedder (2011) describe the first waves of the data that include the measures of stock market beliefs forming the core of our study; these are part of a survey module developed to assess the effects of the financial crisis on household behaviour and well-being. Next to many background variables, we are able to link several other surveys containing data on financial numeracy and knowledge, probability numeracy, and portfolio choices. Table **??** in the Online Appendix contains the exact references for all variables that we use.

Table 1 contains summary statistics of the covariates we use in our main specification. Throughout the paper, we apply the same sampling restrictions, namely observing at least 5 waves of stock market beliefs. The age structure of our sample skews somewhat older than the adult population. Compared with the 2010 Census, our sample includes more individuals aged between 50 and 65 and less under the age of 30. Women are slightly overrepresented, and individuals in our sample are substantially better educated. The fraction of individuals whose highest educational attainment is high school and below is less than half of what it was in the population in 2010.

Our data include answers to several questions that probe subjects' engagement with the stock market. We use a measure of whether subjects participated in the stock market beyond retirement accounts (such as an IRA, 401(k) and similar). They were also asked to self-assess the extent to which they follow and understand the stock market. Table 1 shows that the majority of the respondents in our sample has not engaged much with the stock market. Three quarters do not own stocks outside of their retirement accounts. Less than half of respondents claim they follow the market; only 40% consider themselves to have a good understanding of it. For a subset of respondents, we also have a measure that explicitly tests their knowledge of past returns. Individuals were first asked to select the sign of the return or indicate that they

Table 1: Descriptive Statistics - Individual characteristics

Variable	Observations	Mean	Std. dev.	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$
Age: ≤ 30	3030	0.14				
Age: $\in (30, 50]$		0.33				
Age: $\in (50, 65]$		0.37				
Age: ≥ 65		0.16				
Female	3030	0.59				
Education: High school or less	3030	0.18				
Education: Some college		0.38				
Education: Bachelor degree		0.26				
Education: Advanced degree		0.18				
Owns stocks	3030	0.27				
Follows stock market	3010	0.46				
Understands stock market	3010	0.40				
Knowledge of returns: False Sign	2067	0.20				
Knowledge of returns: Don't Know		0.31				
Knowledge of returns: Magnitude too large		0.07				
Knowledge of returns: Correct		0.42				
Financial Numeracy	1564	0.82	0.22	0.52	0.86	1
Financial Knowledge	1564	0.78	0.24	0.46	0.87	1
Probability Numeracy	1940	0.67	0.2	0.4	0.7	0.89

– The observations summarised in the table are restricted to individuals in our final sample.
– For dummy variables, only means are shown.
– Age is set to the within-person median across surveys.
– Education is set to the within-person mode across surveys.
– "Owns stocks" is the within-person mean of a dummy equalling 1 if respondents indicated that their liquid portfolio included stocks or mutual funds. This excludes stock holdings as part of an IRA, 401(k), Keogh or similar retirement accounts.
– "Follows stock market" equals 1 if individuals indicate they follow markets "very closely" or "somewhat" and 0 if "not at all".
– "Understands stock market" equals 1 if individuals rate their understanding of stock markets to be "extremely good", "very good" or "somewhat good" and 0 if they chose "somewhat poor", "very poor" or "extremely poor".
– The categories of "Knowledge of returns" refer to whether respondents were able to recall the return of the Dow Jones over the past year.
– Financial numeracy and knowledge are the first principle component for correct answers, rescaled to lie between 0 and 1, for the two sets of questions in the financial literacy battery referred to as basic and sophisticated in (Lusardi and Mitchell, 2007)
– Probability numeracy is the fraction of correct answers to questions aimed at measuring probabilistic reasoning (Hudomiet, Hurd, and Rohwedder, 2018).

do not know, then the magnitude by choosing one of several bins. As the actual returns were between 7% and 16% when respondents answered the question, we count answers of $[0\%, 10\%]$ and $[10\%, 20\%]$ as correct. 42% of respondents fall into this category. 7% estimate a larger value, 31% choose the "don't know" option and twenty percent give a negative sign.

The ALP data contain a standard battery of questions measuring financial literacy, which is a key predictor of financial decision making (Lusardi and Mitchell (2014)). We use data from a wave that was in the field between March and September 2009. The battery consists of two sets of questions aimed at measuring financial numeracy (often called "basic financial literacy") and financial knowledge ("advanced financial literacy"), respectively (eg Lusardi and Mitchell (2007)). We extract the first principal component from each block of questions and scale each measure to have support between zero and 1. Both measures are left-skewed and have means of 0.82 and 0.78, respectively.

Finally, we use the probability numeracy battery developed in Hudomiet, Hurd, and Rohwedder (2018), who find that few people understand complex laws of probability but that most people have a basic understanding. We limit ourselves to a basic measure by using the fraction of correct answers across questions an individual answered. Table 1 shows that the average fraction of correct responses is 0.67 with a standard deviation of 0.20, implying substantial variation in probability numeracy.

2.1 Measures of stock market beliefs

The data on stock market beliefs stem from the survey module "Effects of the Financial Crisis" (Hurd and Rohwedder, 2011), which was fielded between late 2008 and early 2016 with a total of 61 waves. The first two waves were collected in November 2008 and March 2009. Starting in May 2009, data were collected monthly until April 2013. Afterwards, the surveys ran at a quarterly frequency until they ended in January 2016. As we are interested in belief formation, we restrict ourselves to individuals who responded at least five times to the belief measures. In total, we have on average 26 waves of data for 3030 individuals for a total of 77310 observations available. Figure **??** in the Online Appendix shows the distribution of survey waves by individual.

The belief measures we analyse consist of three points on the subjective cumulative distribution function. Let p_t be the value of the Dow Jones Industrial Average at time t, and $R_{t\to t+12} := \frac{p_{t+12}-p_t}{p_t}$ the return on the Dow Jones in 12 months. We are very explicit about the notation when it comes to timing because questions about annual returns are asked at a monthly or quarterly frequency, which may lead to confusion otherwise. All time indices in this paper indicate months. For $\Pr(R_{t\to t+12} > 0)$ the question was:

> We are interested in how well you think the economy will do in the future. By next year at this time, what are the chances 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?

For $\Pr(R_{t\to t+12} > 0.2)$ the question was:

> By next year at this time, what is the percent chance that mutual fund shares invested in blue-chip stocks like those in the Dow Jones Industrial Average will have increased in value by more than 20 percent compared to what they are worth today?

For $\Pr(R_{t\to t+12} \leq -0.2)$ the question was:

> By next year at this time, what is the percent chance that mutual fund shares invested in blue-chip stocks like those in the Dow Jones Industrial Average will have fallen in value by more than 20 percent compared to what they are worth today?

From the three points on the cumulative distribution function, we construct an approximation of an individual's expected return to serve as our primary dependent variable. The approximation is as follows:

$$E[R_{t\to t+12}] = \sum_{j}^{4} E[R_{t\to t+12}|R_{t\to t+12} \in I_j] \cdot \Pr(R_{t\to t+12} \in I_j)$$

where the intervals I_j are $[-\infty, -0.2], [-0.2, 0], [0, 0.2]$ and $[0.2, \infty]$. The probabilities in these expressions are observed in the data. We set the conditional means they average to the midpoint of each interval. For the open intervals, we set the lower and upper bounds to the 1st and 99^th percentiles of the historical distribution of the Dow Jones' return (-0.32 and 0.43, respectively). Rather than dropping sets of observations that violate monotonicity of the cumulative distribution function (i.e., $\Pr(R_{t\to t+12} \leq -0.2) \leq \Pr(R_{t\to t+12} \leq 0) \leq \Pr(R_{t\to t+12} \leq 0.2)$), we restore weak monotonicity by setting its values at -0.2 and/or 0.2 to its value at 0. Such monotonicity violations are very common in this question format—for example, around 40% of responses both in the data of Hurd, Rooij, and Winter (2011) and in our own. We will return to inconsistencies in the next section.

In robustness checks, we avoid assumptions on monotonicity violations altogether by focusing on the probability of a positive return (e.g. Dominitz and Manski, 2007, also had even though more more measures available and discarde them presumably for such reasons). Table 2 shows summary statistics for within-person means of the different belief measures, i.e., the mean return and the three points on the cumulative distribution function. We first calculate means for each individual and then average across individuals, thereby weighting every sample participant equally regardless of the number of times she participated. The variation across the different points of the distribution function appears reasonable and all measures exhibit substantial variation across individuals.

Table 2: Individual belief measures averaged over time

	Mean	Std. dev.	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$
$E[R_{t\to t+12}]$	0.5	5.8	-6.9	0.6	8.1
$\Pr(R_{t\to t+12}) > -0.2$	74.6	13.4	55.0	76.5	90.9
$\Pr(R_{t\to t+12}) > 0$	44.0	17.8	19.4	45.3	67.9
$\Pr(R_{t\to t+12}) > 0.2$	26.8	14.2	9.1	25.3	47.1

N = 3030. Units in percentage points.

2.2 Stylised Facts

Our data on stock market beliefs has a number of distinct features that motivate our modelling choices below. Most of these characteristics are similar to those in other data; we briefly highlight them here and provide a full set of statistics in Online Appendix **??**. We do present the co-movements of beliefs with recent information in the main text because these are of particular interest and, when it comes to economic news, novel.

Similar to findings summarised in Hurd (2009), average beliefs are well below historical returns. For example, the mean of individuals' expected returns in our data is 0.5%, compared to a historical value of 7.3%. Beliefs do not only vary across individuals as shown in Table 2, but also within individuals over time. The magnitude of within-variation is similar to the magnitude of between-variation. Regression analyses controlling for many other factors show that beliefs of financially sophisticated and knowledgeable individuals are more optimistic. They also reveal that their beliefs are more likely to constitute actual probability judgements in two different senses.

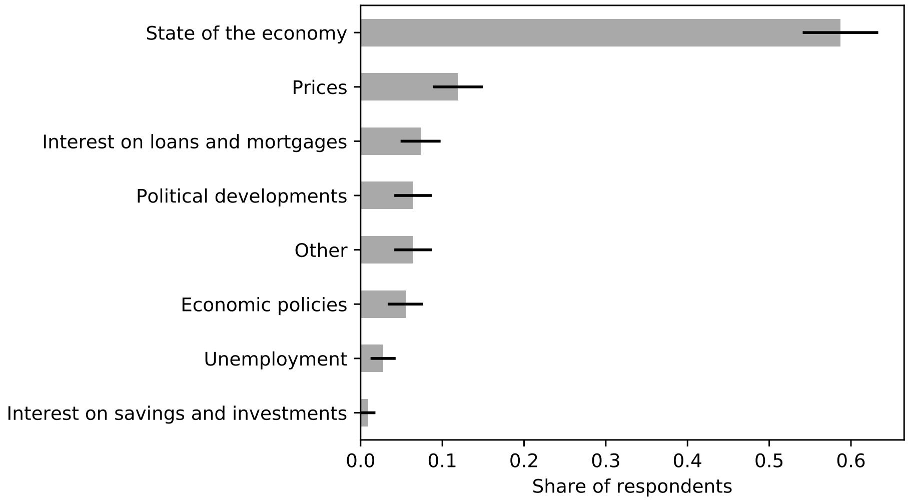
First, Bruin et al. (2000) argue that 50% answers might indicate that individuals are epistemically uncertain about an event rather than expressing subjective beliefs of equal likelihoods. Following up on that observation, the questionnaires that we use confront respondents who gave an answer equal to 50% for $\Pr(R_{t\to t+12} \leq 0)$ with a follow up question. It asks them to clarify whether they mean that the Dow Jones is equally likely to rise as it is to fall, or whether they want to express that they are unsure what to do (also see Enke and Graeber, 2019). 53% of all answers when the follow up question was encountered turn out to be best characterised as expressing uncertainty that way. Second, if respondents are unsure about the behaviour of the Dow Jones index, they will be more likely to give sets of answers that violate monotonicity. Regressions reported in Appendix **??** show that even after controlling for numerous other characteristics, measures of stock market following, financial numeracy, and financial knowledge are associated with substantially lower rates of monotonicity violations.

A recent literature has documented that average return expectations covary with recent stock market movements. Kezdi and Willis (2008) and Hurd (2009) noted this phenomenon early on. Greenwood and Shleifer (2014) find evidence for it across a variety of data sets; they also coined the term "extrapolative expectations". We corroborate this finding. In addition, we find evidence that individuals, on average, react to other types of information. In a small-scale ALP survey that overlaps with individuals in our main data, respondents were first asked about the probability of a stock market gain, much in the same way as the first question reproduced in Section 2.1.[1] After a short interlude of questions not of interest to us, they were asked to state what they most thought about when answering this question. Figure 1

[1]The precise question was "By next year at this time, what is the percent 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?"

shows the distribution of possible answers; the state of economy is by far the most common answer.

Figure 1: What respondents think most about when contemplating future stock prices

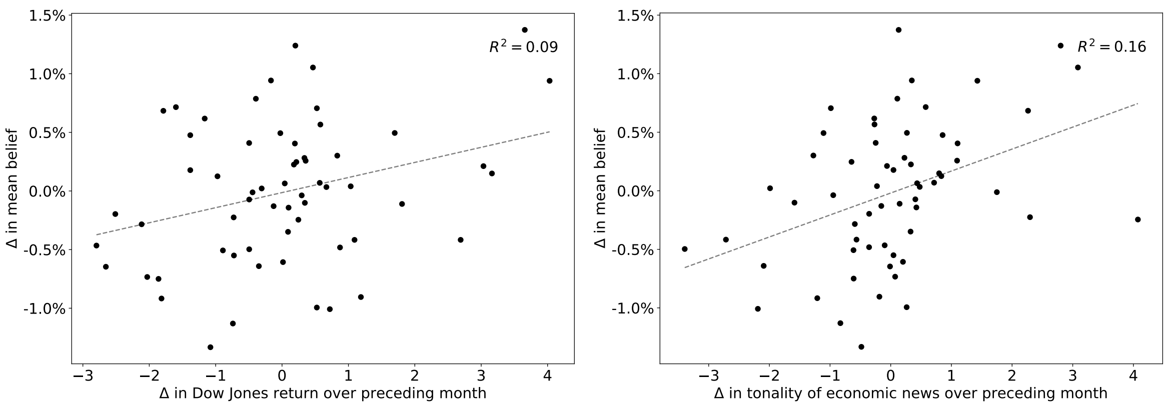


Only includes individuals who overlap with our main sample, $N = 114$.

This finding and the fact that only 42 percent of individuals in our sample have reasonable knowledge of how the Dow Jones changed over the preceding year (see Table 1) lead us to include additional information that subjects may use to form beliefs about stock returns. We hence obtained data on the tonality of economic news on major TV networks. We construct our measure using data provided by Media Tenor International, who had analysts classify evening news segments on CBS, Fox, and NBC in terms of what they refer to and whether the news is positive, neutral or negative. We take all news items referring to the state of the economy on day d and score positive items (pos) with 1, neutral items (neu) with 0 and negative times (neg) with - 1. We define our measure of the tonality of economic news as the average monthly score: $N_{t-1\to t} := \frac{\sum_{d\in[t-1,t]} 1\cdot\text{pos}_d+0\cdot\text{neu}_d-1\cdot\text{neg}_d}{\sum_{d\in[t-1,t]}\text{pos}_d+\text{neu}_d+\text{neg}_d}$.

We investigate the extent to which individuals extrapolate good and bad news, in form of recent stock returns and media reports on the economy as follows. We average expected returns across individuals for every survey wave, take first differences and plot them against the first differences of the Dow Jones Index return over the past month and the first differences of the average monthly news score. As shown in the first panel of Figure 2, an increase in the Dow Jones' returns over the past month of 4.3 percentage points (one standard deviation of the monthly return over our analysis period) is associated with a 0.13 percentage point higher expectation on the return over the next year. In the second panel depicting how first differences in expected returns vary with first differences in economic news, we see a similar pattern. A one unit increase in the news measure corresponds to an increase of 0.19 percentage points in expected returns. To put these numbers into context, Greenwood and Shleifer (2014) find that an increase in the annual return of 20 percentage points (one standard deviation of the annual return over the period on which their regression is based) increases the Michigan Survey expectations 0.78 percentage points [2].

Figure 2: Average expected returns extrapolate stock prices and follow the tone of news



Both panels depict survey-to-survey changes in means of expected returns on the y-axis. The x-axis depict survey-to-survey changes in the standardised measures of recent monthly returns and news, respectively. X-axis units in time series standard deviations.

2.3 Empirical strategy

The stylised facts about individuals' stock market beliefs have shown that beliefs are very heterogeneous within and across individuals; that part of the between-variation is explained by financial sophistication; that the beliefs' evolution over time covaries with past returns and news about the economy; and that measures of beliefs vary in their informational content about true beliefs, which again varies systematically with financial sophistication. Together,

[2]Greenwood and Shleifer (2014) obtain these results by regressing expected returns on annual returns. Figure 2 depicts the first differenced version of that regression, replacing annual with monthly returns, and separately also for news

these facts point towards putting between-person heterogeneity at the centre of a model of beliefs and their evolution over time. In particular, models that treat heterogeneity as an incidental parameters problem—fixed effects estimation being arguably the most prominent example—are doomed to fail. We expect individuals to differ in their levels of beliefs, in their belief volatility over time, in how they change their beliefs in response to information about recent returns and economic news, and in the extent to which the measures we have at our disposal represent actual, accurate beliefs. At the same time, we need to impose some restrictions across individuals because our panel dimension is too short to allow for estimating models at the level of the individual.

We thus assume that we can summarise heterogeneity in belief formation processes by using a discrete set of groups. As long as the number of groups does not become too large, it allows us to describe the multidimensional patterns of heterogeneity in an accessible way; this would be difficult for many continuous distributions. Our main specification for belief formation is a linear model of the form:

$$E[R_{t\to t+12}]_{i,t} = \alpha_g + \sum_{l=0}^{L} \left(\beta_{g,l} R_{t-1-l\to t-l} + \gamma_{g,l} N_{t-1-l\to t-l}\right) + u_{i,t}. \tag{1}$$

We take $u_{i,t}$ to be independently and identically distributed across individuals and over time. We assume that all heterogeneity beyond that is captured by the coefficients. Put differently, we assume that there is a discrete number of groups G. All parameters of the model are allowed to differ at the group level, indexed by g: The intercept α_g measures the persistent degree of optimism or pessimism, the parameters $\beta_{g,l}$ measure how returns l months ago influence current beliefs, and $\gamma_{g,l}$ do the same for economic news N.

We estimate the model for $L = 0$, i.e., using only the most recent returns and news, and for $L = 6$. The latter allows us explore potential patterns of momentum in beliefs. We also experimented with averages across longer periods—e.g., much of the literature has considered annual returns—but found monthly intervals to provide the best fit. When constructing R and N, we are exact to the day on which individuals completed the survey.

In order to estimate the model, we employ the two-step method of Bonhomme, Lamadon, and Manresa (2017). In the first step, we classify individuals into a discrete set of G groups using moments of the both dependent and explanatory variables. In the second step, we estimate the coefficients in (1) separately for each group. This method is computationally simple and very transparent, providing easily interpretable groups.

Following Bonhomme, Lamadon, and Manresa (2017), we use the k-means algorithm in order to classify individuals into groups. The algorithm works by choosing the group assignments that minimise the sum of squared deviations between included variables and the group-wise means of these variables. The problem is NP-hard, but a number of heuristic algorithms exist that work well in practice. The method is widely used in machine learning; we use the implementation in the Python library *scikit-learn* (Pedregosa et al., 2011). Since solutions to the k-means objective are sensitive to the scaling of variables, we follow common practice and standardise each classification variable to have mean zero and unit variance in the cross-section of individuals.

In order to classify individuals into groups, we use moments of their stated beliefs and their relation with the explanatory variables. In particular, for each individual series of $\Pr(R_{t\to t+12} > -0.2)_{i,t}$, $\Pr(R_{t\to t+12} > 0)_{i,t}$, and $\Pr(R_{t\to t+12} > 0.2)_{i,t}$, we use its mean, its standard deviation, and its covariances with the return of the DJ as well as economic news, each measured over the month before the survey. These capture the dimensions level, volatility and response to information. In addition, we use the fraction of beliefs satisfying strict monotonicity, and the fraction of beliefs for which respondents did not indicate that beliefs expressed that they were unsure (or were not given the chance to do so). These capture the dimension of epistemic uncertainty. This makes for a total of fourteen time-constant moments that vary across individuals. We make this choice for two reasons. First, these moments exclusively use raw data and make no additional assumptions. This contrasts with, for example, expected returns, which entail a number of assumptions as detailed in Section 2.1. Second and more importantly, these are the key moments that should be informative on group-level heterogeneity along the dimensions we are interested in, as required for the analysis in Bonhomme, Lamadon, and Manresa (2017).[3]

[3]Note, however, the conceptual difference in that we assume that there is a discrete number of groups whereas the focus of the theoretical analysis in Bonhomme, Lamadon, and Manresa (2017) is on controlling for continuous unobservables.

Table 3: Moments and corresponding dimension

Moments	Dimensions
Mean probability that $R_{t,t\to t+12} \in (-0.2, \infty)$	Level
Mean probability that $R_{t,t\to t+12} \in (0, \infty)$	Level
Mean probability that $R_{t,t\to t+12} \in (0.2, \infty)$	Level
St. dev. of prob. that $R_{t,t\to t+12} \in (-0.2, \infty)$	Volatility
St. dev. of prob. that $R_{t,t\to t+12} \in (0, \infty)$	Volatility
St. dev. of prob. that $R_{t,t\to t+12} \in (0.2, \infty)$	Volatility
Cov. of prob. that $R_{t,t\to t+12} \in (-0.2, \infty)$ and returns	Response to Information
Cov. of prob. that $R_{t,t\to t+12} \in (0, \infty)$ and returns	Response to Information
Cov. of prob. that $R_{t,t\to t+12} \in (0.2, \infty)$ and returns	Response to Information
Cov. of prob. that $R_{t,t\to t+12} \in (-0.2, \infty)$ and news	Response to Information
Cov. of prob. that $R_{t,t\to t+12} \in (0, \infty)$ and news	Response to Information
Cov. of prob. that $R_{t,t\to t+12} \in (0.2, \infty)$ and news	Response to Information
Fraction of beliefs satisfying strict monotonicity	Epistemic Uncertainty
Fraction of beliefs expressing probability judgements	Epistemic Uncertainty

3 Results

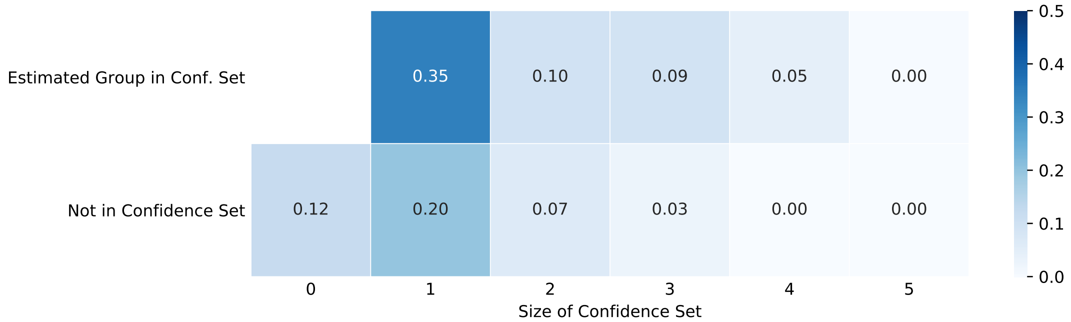
We first outline describe the classification into groups, including a diagnostic test. We then describe our main results before reporting on a number of robustness checks. The last part of this section explores the extent to which the groups we identify can be described by having rational expectations.

3.1 Classification into groups

In our main specification, we use five groups because this was the minimum number of groups where no economically meaningful intergroup differences were blurred. Larger numbers led to little additional economic insights and eventually to apparent overfitting. We will be more precise on this below in Sections 3.2 and particularly in 3.4, where we also consider alternative choices for the number of groups. As noted before, we require five belief measures per individual and use the moments listed in Table 3 as an input to the k-means algorithm.

Dzemski and Okui (2018) have developed a diagnostic test for clustering methods such as our classification step. Their procedure yields a unit-wise confidence set of group membership for each individual. It is constructed by testing the null hypothesis that individual i's true group

Figure 3: Unit-wise 10% confidence sets by size and inclusion of estimated group



The numbers refer to the share of individuals in each cell.

g_i^0 is g for all groups $1, \ldots, G$. The elements of the confidence sets are then those groups for which the null hypothesis cannot be rejected for a pre-specified confidence level. The test is based on the insight that if $g_i^0 = g$, then $E[(y_{i,t} - x_{i,t}^T \theta_g)^2)] \leq E[(y_{i,t} - x_{i,t}^T \theta_h)^2)]$ for all possible groups h (collecting all model parameters in the vector θ).

First of all, it is important to note that all group sizes are substantial. The largest group's share is 26% and that of the smallest is 13%. Figure 3 shows the distribution of unit-wise 90% confidence sets by their size and by whether they contain the estimated group. With 35% of individuals, the estimated group assignment being the only element in the set is the most common occurrence. For another twenty-three percent, the estimated group is in the confident set, but in addition to other groups. So for almost 60%, the estimated group is in the confident set. At the same time, very few confidence sets have more than three elements. Given that we have rather noisy data (compared to, say, the classification of states or countries, as the examples in Dzemski and Okui, 2018), these results demonstrate that our approach yields reasonable results even for a relatively low number of groups.

Nevertheless, a sizable fraction of confidence sets do not include the estimated group. Part of this is a reflection of the fact that the test is based on goodness of fit of our model (1), wheras the k-means procedure gives equal weight to all included features. Most notably, one would not expect the standard deviation over time to improve the fit of model (1). Indeed, the next section will demonstrate that level differences in expectations are the dominant component for improving model fit. Insofar as the k-means algorithm compromises splitting individuals along their expectation level to accommodate splitting them along differences in belief dispersion and association with returns or news, this will trivially result in a larger share being assigned to groups that are not in their confidence set than if one was assigning groups based on similar goodness-of-fit criteria as the test uses (as, for example, in the clustering method of Bonhomme and Manresa, 2015). The reason confidence sets can be empty is similar; none of the estimated groups provide a good fit for some individuals whose expectation level is far from that of any of the groups.

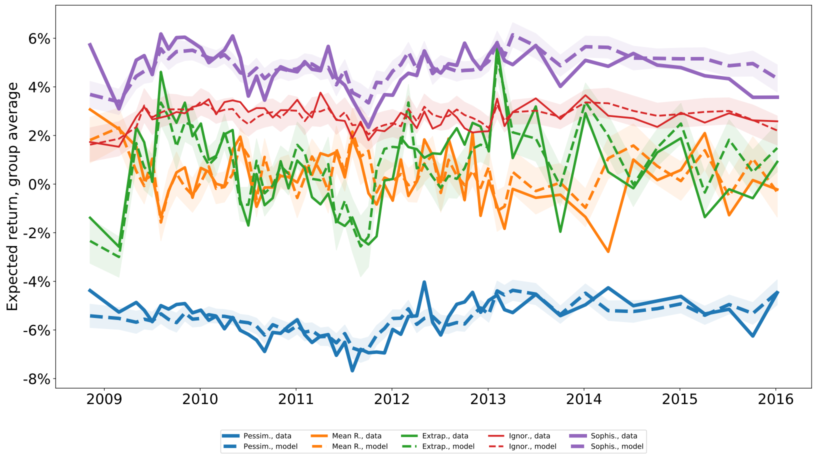
3.2 Heterogeneity in groups' behaviour and characteristics

We order the 5 groups by their average expected returns and refer to them as pessimists, mean reverters, extrapolators, ignorants, and sophisticates, respectively. Each group's label captures the characteristic of the moments used for classification that makes it stand out from the others the most. The key results are summarised in three figures and two tables. Figure 4 plots the data averages (solid lines) versus the model predictions (dashed lines) of expected returns over time. Table 4 summarises the group means and standard deviations of expected returns, averaged within each group and survey wave, over our sample period. Figure 5 plots the reaction of groups to changes in past returns and news, respectively. Table 5 shows prevalence of monotonicity violations and the fraction of answers expressing epistemic uncertainty, respectively. Finally, Figure 6 presents the mean values of various covariates for each group.

Before describing each group in turn, we note that differences across all dimensions are important. The levels of beliefs in Figure 4 are strikingly different and—except for the mean reverters and extrapolators—hardly ever cross. The volatility over time is largest for mean reverters and extrapolators; it is by far smallest for ignorants with the other two groups in between. The reactions to both stock prices and news depicted in Figure 5 are substantial and very different.

Pessimists (25% of individuals) consistently expect the return of the Dow Jones to be negative and substantially so (-5.6%). Their beliefs do not vary too much over time, although they seem to be a bit more optimistic in the second half of our sample period (but still far below any other group). This seems to be due to better economic news in this period, to which they respond positively. Their beliefs do not react to past return. Along the dimensions of

Figure 4: Data vs. predicted expected return of the Dow Jones index, by group



The solid and dashed lines are within survey and group means of individual data points and model predictions. Shaded regions are within survey × group means of individual 95% confidence intervals for the estimated regression function. Line widths are proportional to group sizes.

knowledge and numeracy, pessimists appear to be in the middle of the distribution along with mean reverters and extrapolators.[4]

Table 4: Long run moments by group

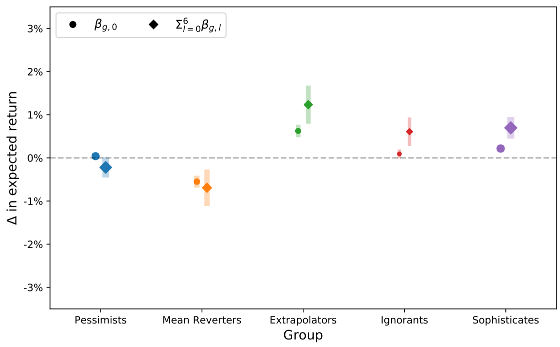
	Data		Model	
	Mean	St. Dev.	Mean	St. Dev.
Pessimists	-5.57	0.80	-5.58	0.40
Mean Reverters	0.29	1.10	0.37	0.75
Extrapolators	0.79	1.81	0.91	1.43
Ignorants	2.78	0.49	2.80	0.15
Sophisticates	4.74	0.86	4.83	0.38

N = 3030. Units in percentage points. Expected returns are averaged within each group and survey wave, mean and standard deviation are calculated over the resulting time series points.

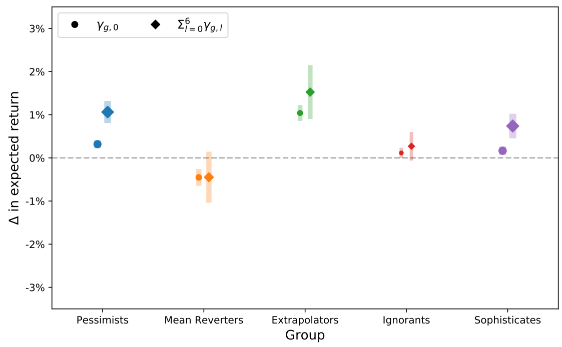
[4]The fractions of monotonicity errors and of beliefs expressing epistemic uncertainty (Table 5) seemingly stand in contrast to this, but they are probably due to somewhat mechanical effects. For monotonicity violations, giving low answers to $\Pr(R_{t\to t+12} > 0)$, as pessimist frequently do, will c.p. lead to less monotonicity errors if stated beliefs are subject to survey response error. This is because when stating the last elicited belief, $\Pr(R_{t\to t+12} < -0.2)$, the margin for avoiding a monotonicity error is larger when $\Pr(R_{t\to t+12} > 0)$ was small. In line with this explanation, the gap in monotonicity violations between pessimists and mean reverters / extrapolators is largely driven by violations of $\Pr(R_{t\to t+12} < -0.2) \leq Pr(R_{t\to t+12} < 0)$. In order to arrive at the follow-up question on epistemic uncertainty, an individual needs to use 50% when asked about the chance the Dow will increase. Pessimists feature very few 50% answers.

Figure 5: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Past returns of the Dow Jones



(b) Past tonality of economic news



Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones (Panel a) and in the most recent tonality of economic news over one month (Panel b). Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones and the preceding tonalities of economic news, respectively. Shaded lines show the width of 95% confidence intervals. Marker and line widths are proportional to group sizes.

Mean Reverters and Extrapolators (19% and 17% of individuals, respectively) are also rather pessimistic, expecting a return of about zero. Individuals in these two groups are similar in observable characteristics. Their key difference, and reason for the labels we chose for them, can be seen in Figure 5: Extrapolators expect recent trends to continue and do so more than any other group. Mean reverters follow the opposite pattern: They become less optimistic following a good performance of the Dow Jones or positive economic news, and are the only group which reacts in this way. Hence, the lines in Figure 4 frequently cross and move in opposite directions survey to survey.

The fact that mean reverters and extrapolators are very similar in terms of observable characteristics, but react in completely different ways to information, underlines the importance of classifying individuals in terms of features related to their stated beliefs. Considering only observed heterogeneity, as in classical regression analysis (see Section **??** of the Online Appendix), would necessarily hide this important dimension of behaviour.

Ignorants (13% of individuals) are seemingly the second most optimistic group. Their average belief that the Dow Jones will increase is almost exactly 50% and they expect a return of 2.8%. Compared to the other groups, ignorants are notable for their very low belief variability. Panel B of Figure 6 shows that their average is near the tenth overall percentile and Figure 4

Table 5: Measures of epistemic uncertainty by group

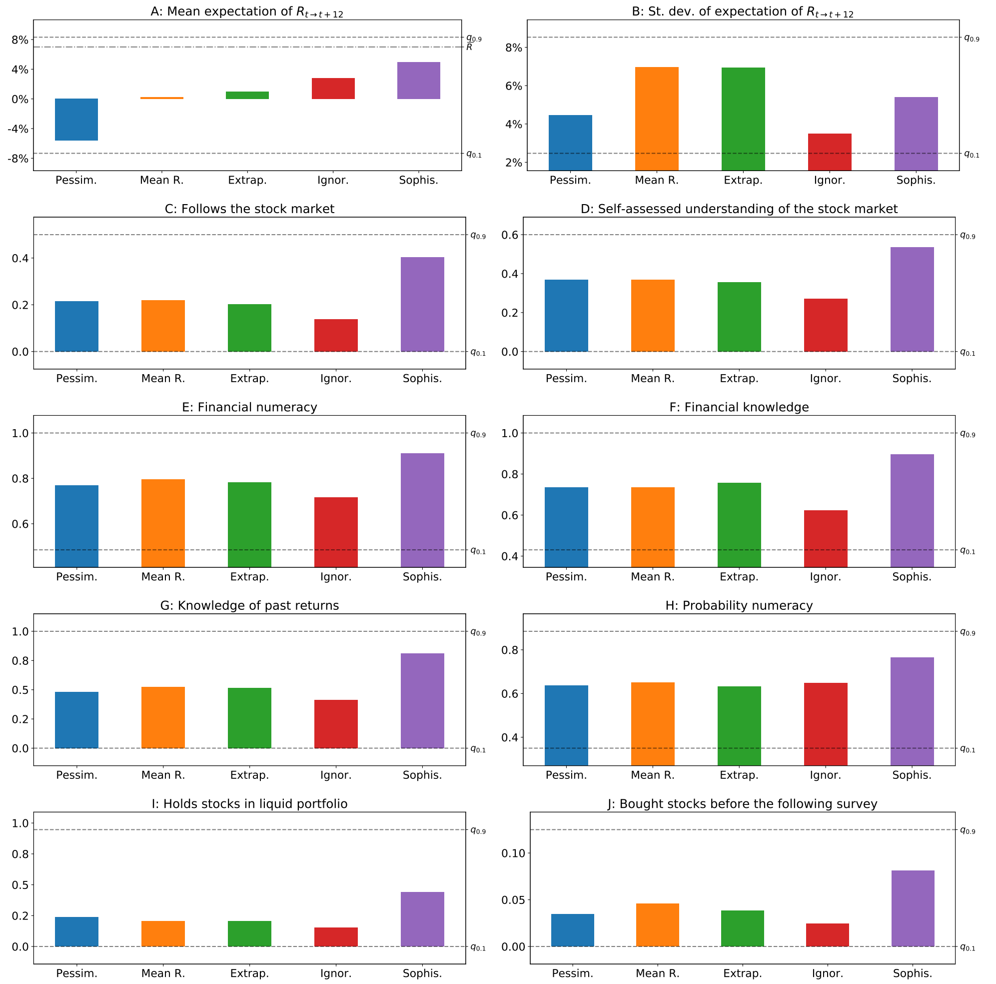
	Fraction of belief sets satisfying strict montonicity	Fraction of beliefs expressing subjective probabilities
Pessimists	0.67	0.96
Mean Reverters	0.42	0.90
Extrapolators	0.41	0.88
Ignorants	0.14	0.61
Sophisticates	0.75	0.95

N = 3030. A belief set satisfies strict monotonicity if $\Pr(R \leq -0.2) < \Pr(R \leq 0) < \Pr(R \leq 0.2)$. Beliefs express subjective probabilities if, for the question asking about the probability of an increase of the DJ, the belief is not 0.5, or it is and in the follow up question, the respondent indicated this means an equal likelihood.

visualises how comparatively little individuals belonging to this group change their beliefs. As Figure 5 shows, their beliefs also covary least with returns and news. In addition to the belief that the Dow Jones will increase, the other two subjective beliefs are, on average, close to 50% as well which is incompatible with strong monotonicity of the cdf. Ignorants are most likely out of all groups to violate monotonicity, with only 10% of belief sets satisfying it. One key reason for that is that where other groups express subjective probabilities with their stated beliefs 90% of the time and more, this is barely more than 60% for ignorants, which is below the tenth overall percentile. In other words, they use 50% answers to express epistemic uncertainty about stock returns. In line with their apparent lack of informedness, ignorants also have the lowest scores when it comes to following and understanding the stock market, knowledge of past returns and financial knowledge. Though seemingly more optimistic than other groups, all our indicators suggest that the stated beliefs of these individuals are limited in terms carrying quantitative information, and need to be interpreted with caution.

Sophisticates (26% of individuals), the most optimistic group, expect the Dow Jones to yield an average return of 5%. That number is relatively close to the historical performance of 7.3%. In addition to having beliefs that are most accurate compared to the historical distribution, sophisticates also stands out from the others in terms of experience with the stock market and knowledge relating to it. They are more likely to describe themselves as following and understanding the stock market, they have a superior knowledge of historical returns and greater financial knowledge. Sophisticates have the best understanding of probability calculus, are least likely to express beliefs that violate monotonicity of the cumulative distribution function (more than 80% of their belief sets satisfy strict monotonicity), and, together with pessimists, they use beliefs to express subjective probabilities most often.

Figure 6: Observable characteristics by unobserved heterogeneity group



N = 3030, smaller for some panels depending on the availability of covariates, see Table 1. Bars show group means, dashed lines are the bottom and top decile with respect to the individuals of all groups taken together. *Variable definitions:*: Financial numeracy and knowledge: First principle components loading on variables indicating whether a respondent correctly answered numerical and knowledge based questions, scaled to the unit interval; Probability numeracy: Fraction of correct answers to questions about probability theory; Knowledge of past returns: False sign (0), don't know (⅓), magnitude too large (⅔), sign and magnitude correct (1); Understanding of the stock market: Extremely bad (0), very bad (⅕), bad (⅖), good (⅗), very good (⅘), extremely good (1); Follows stock market: Not at all (0), somewhat (½), closely (1).

3.3 Stock ownership and trading behaviour

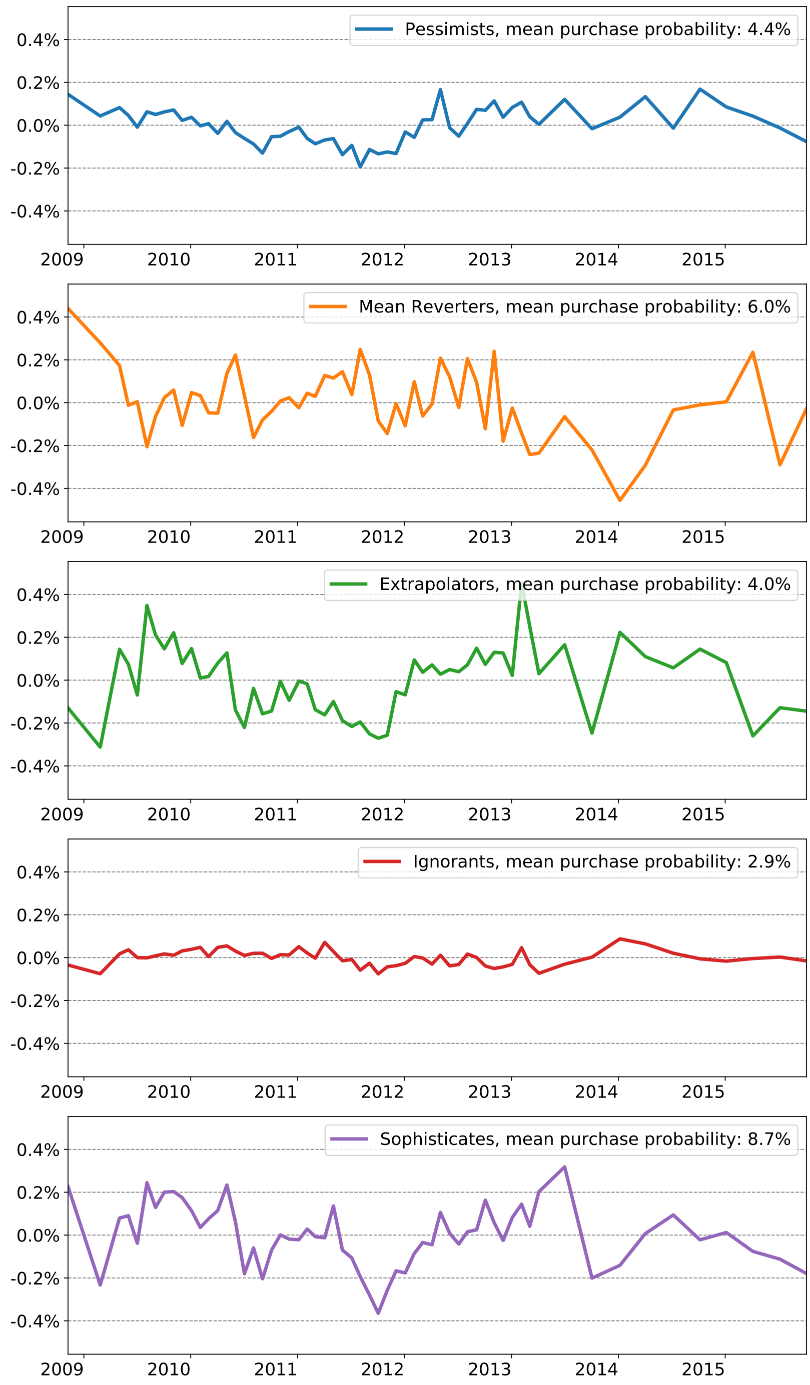
The differences in beliefs and their trajectories translate into very different behaviour when it comes to portfolio choice. Panel I of Figure 6 shows that stockholding is lowest for ignorants (15%) and highest for sophisticates (44%), with the other groups below 20%, too. Trading behaviour follows a similar pattern (Panel J).

In order to further investigate this, we run a Probit regression of buying stocks in the subsequent period on a set of group fixed effects and the return expectation at the time of the survey. Because of the low baseline probabilities, it is important to use a nonlinear model as opposed to a linear probability model. This means that controlling for fixed effects is infeasible due to the incidental parameters problem. The average partial effects of increasing expectations by one group-level standard deviation are 0.14% for sophisticates and 0.03% for ignorants, respectively. For extrapolators, due to the comparatively greater volatility of their expectations (see Table 4) the effects are 0.16%. The other groups are somewhere in between.

Figure 7 shows that these patterns translate into very different predicted purchasing patterns over time. Pessimists and, even more so, ignorants hardly change their behaviour over time. Their predicted purchasing probabilities fluctuate slightly around low average values. The other three groups show much more differences over time. Again, this often goes in opposite directions. Not surprising, mean reverters would have higher than average values during the in the aftermath of the financial crisis, a time when the tone of news was also dire. Extrapolators show the opposite pattern. Sophisticates have the highest trading probability and a variability that is slightly below that of mean reverters or extrapolators.

It is once more important to note that these rich patterns of heterogeneous decision-making only surface because of our classification into groups based on (time-series) features of the dependent variable. Controlling for observed characteristics could only induce vertical shifts in trading behaviour, but no reversal of patterns. These are important, however, to generate potential trade between groups. Again, our findings mirror those reported in Giglio et al. (2019), who find that beliefs to be reflected in portfolio allocations and a small but predictable effect of belief changes on trading patterns.

Figure 7: Predicted probability of buying stocks, group averages over time



N · T = 70114, N = 3029. Group and survey average predicted probabilities from a probit regression of a stock buying indicator in the next period, provided that is within 120 days, on group indicators and expected returns.

3.4 Discussion and robustness of results

Our preferred model explains more than a quarter of the variation in expected returns (see Table **??**). This differs by a factor of one hundred from the same model without unobserved heterogeneity. Similarly, a standard regression model with lots of observed heterogeneity can explain only 12%.[5] This squares well with Giglio et al. (2019), who find as one of their five facts that "Beliefs are mostly characterized by large and persistent individual heterogeneity; demographic characteristics struggle to explain why some individuals are optimistic and some are pessimistic." Our analysis underlines their finding and goes beyond it in documenting different belief formation processes.

In Section **??** of the Online Appendix, we relax the requirement of observing at least five sets of belief measures per individual to a minimum of three. The broad pattern of groups remains the same and we can essentially leave the labels in place (mean-reverters and extrapolators switch their places in terms of average expected returns and the latter group shrinks by about one third). Similarly, the group assignments remain very stable when requiring a minimum of fifteen periods per individual. Note that this ensures that the version of (1) with $L = 6$ lags would be identified individual-by-individual. The results are presented in Section **??** of the Online Appendix. 86% of the respondents that meet the stricter requirement are assigned to the same group as before. The number is lowest for sophisticates at 73% (see Table **??**), most of the remainder is assigned to the group of extrapolators.

As detailed in Section 2.1, our measure of expected returns makes a number of assumptions. In Section **??**, we thus report results on a specification that uses the raw data on the probability of a stock market gain as the dependent variable. This also makes the analysis comparable to Dominitz and Manski (2011). By construction, the distribution of groups is exactly the same as in our main model (some tables and graphs shown in the other cases are thus superfluous) and, reassuringly, the diagnostic tests looks very similar, too. The time series look very similar to before with four clearly distinguishable levels; mean reverters and extrapolators are again on a similar level, crossing frequently. The reactions to simulated shocks show a similar

[5] See Table **??** for the precise results. The first regression is a linear probability model of the level of beliefs on the past six months' returns and news. The "kitchen-sink"-approach additional includes a quadratic in age, sex, education in four categories, ethnicity in five categories, 18 levels of household income, various measures of stock market experience and knowledge, probability and financial numeracy.

pattern to Figure 5, if anything, lags seem to have slightly stronger effects towards building up momentum.

Sections **??**, **??**, **??**, and **??** show the results for 3, 4, 7, and 15 groups, respectively. The results further motivate our choice of G. With three groups, the distinction between extrapolators and mean reverters is blurred with both mostly being allocated to the second group, resulting in extrapolation on average. In the case of four groups, the first four groups remain very stable (each retains more at least 94% of its previous members), but the group of sophisticates is distributed across the other groups with two thirds being pooled with pessimists (see Table **??**). Based on those results, one may conclude that the most optimistic group was mostly made up by respondents with a severe lack of understanding or interest. It also blurs the features of the other groups; most notably, the average expectations of pessimists go up by two percentage points.

Moving from five to six or seven groups has effects almost exclusively for the groups of mean reverters, extrapolators, and ignorants. Both other groups retain more than three quarters of their members and all their characteristics remain very stable. Some of the clusters become fairly small and relative to the other lines in Figure **??**, their data averages and predictions are very unstable. Consequently, the patterns become stronger, particularly on the extrapolation side. The positive interpretation of these patterns would be that some groups of individuals are reacting very strongly to current trends indeed; a sceptic may think that we are fitting noise in the data. In any case, we do not believe that one gains much additional insights from this relative to the case with five groups. The main reason for showing the results for fifteen groups is to demonstrate that while feasible, the algorithm clearly starts fitting noise. For example, group 12 consists only of 22 individuals. Note that the diagnostic test described in Section 3.1 becomes computationally infeasible.

An important feature of many probabilistic statements is rounding. Dominitz and Manski (1997) document a number of facts, including that most non-extreme probabilities are reported in multiples of five. Manski and Molinari (2010) provide a partial identification analysis and conclude that inference should ideally rely on weaker assumptions than the usual practice of ignoring rounding. Kleinjans and Soest (2014) develop an approach for panel data, which classifies individuals into rounding types. That is, the extent of rounding in a given context is seen as a time-constant behavioural trait of an individual. Heiss et al. (2019) apply this model in a similar context to ours. Note that this model is hard to square with a fact noted in Dominitz and Manski (1997), namely that rounding to multiples of 1 is much more prevalent in the tails of the distribution. This is also the case in our data, see Section **??** of the Online Appendix. Our main analysis incorporates a type of rounding behaviour, namely to 50%. This is similar in spirit to some of the analyses in Manski and Molinari (2010), who also use follow-up questions to explore the nature of 50% responses. As we see lots of rounding to multiples of 10%, in a robustness check we add indicators to Equation (1) of whether any of the three subjective probabilities involved in the construction of expected returns is divisible by 10%. The coefficients of our main model are largely unaffected.

3.5 Rational expectations tests

Greenwood and Shleifer (2014) and Giglio et al. (2019) are just two examples of a large literature challenging the rational expectations paradigm for the average investor. In the light of our focus on heterogeneous belief formation processes, it seems very natural to ask whether some groups' belief formation processes may be consistent with rational expectations. In order to do so, we treat expectations as forecasts and analyse the predictability of forecast errors. We apply the methodology of Coibion and Gorodnichenko (2012, 2015), which yields a direct test of whether expectations are rational. In particular, forecast errors of full information rational expectations should be unpredictable with any information I_t at time t because they equal the true expected value of the variable to be forecasted given the information: $E\Big[R_{t\to t+12} - E[R_{t\to t+12}\,|\,I_t]\,|\,I_t\Big] = 0$. Non-full information rational expectation forecast errors should be unpredictable with any information in a forecaster's information set, though they might be with information the forecaster is not aware of or does not use. This insight allows for testing the rationality of expectations without knowing too much about either the true data generating process or what information forecasters use.

We follow the methodology of Coibion and Gorodnichenko (2015) who specify the information set I_t to be the forecast revision. Let $F_tR_{t\to t+h}$ be the forecast of the return $R_{t\to t+h}$ at time t of an individual. Forecast errors are then defined as $FE_t := R_{t\to t+h} - F_tR_{t\to t+h}$ and forecast revisions as $FR_t := F_tR_{t\to t+h} - F_{t-1}R_{t\to t+h}$. Regressing forecast errors on forecast revisions then tests the rationality of expectations, and the sign of the slope coefficient measures whether expectations overreact or underreact to information. If expectations are rational the slope coefficient is zero. A negative sign for the slope means an upwards revised forecast is typically followed by a downwards swing in the forecast error. As the regression includes an intercept, this means that the forecast overshoots, its upwards adjustments went too far. This is overreaction. The logic is reversed for a positive sign, which indicates underreaction.

To estimate this regression with our data, we have to make an assumption. Forecast revisions are defined as the difference of two forecasts of the return $R_{t\to t+12}$; this month's forecast $F_t R_{t\to t+12}$, for which we take individual expected returns, and last month's forecast $F_{t-1}R_{t\to t+12}$. We do not have a direct measure of the latter because beliefs were always elicited about one-year-ahead returns. To proceed, we assume that $F_{t-1}R_{t\to t+12} = F_{t-1}R_{t-1\to t-1+12}$. Hence we assume that last month's forecast of the return a year from then is also how respondents would have answered questions of the form: "What are the chances that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth in thirteen months than what they will be worth in one month?". How strong is this assumption? Writing $R_{t-1\to t-1+12} - 1 = \frac{\frac{p_t}{p_{t-1}}}{\frac{p_{t+12}}{p_{t+11}}} \cdot (R_{t\to t+12} - 1)$, we see that it only depends on the next and last months. If individuals expect the same percentage change in stock prices over the next month as they do from 11 months ahead to 12 months ahead, the assumption is satisfied.

With this assumption, we can write the model as follows:

$$\begin{aligned} FE_{i,t} &= \tau_g + \delta_g FR_{i,t} + \epsilon_{i,t} \\ R_{t\to t+12} - E[R_{t\to t+12}]_{i,t} &= \tau_g + \delta_g \left(E[R_{t\to t+12}]_{i,t} - E[R_{t-1\to t+11}]_{i,t-1}\right) + \epsilon_{i,t} \end{aligned} \tag{2}$$

As before, we allow model coefficients to vary by group. Table 6 contains the results, restricting our sample to consecutive observations during the period where the survey was fielded monthly. Table **??** in the Online Appendix repeats the exercise for our entire sample with very similar results.

As can be seen from the table, all groups overreact with a slope coefficient close to -0.5. This is exactly what we would find if time variation in expectations is uncorrelated with

Table 6: Predictability of forecast errors with forecast revisions

	Pooled OLS	Pooled OLS w groups
Forecast Revision	-0.52	
	(0.01)	
Forecast Revision, Pessimists		-0.52
		(0.01)
Forecast Revision, Mean Reverters		-0.51
		(0.02)
Forecast Revision, Extrapolators		-0.53
		(0.02)
Forecast Revision, Ignorants		-0.50
		(0.01)
Forecast Revision, Sophisticates		-0.53
		(0.02)
R^2	0.12	0.28
N · T	50532	50532

N = 2834. Observations that are not consecutive during the monthly phase of the survey waves are dropped. OLS estimates. Standard errors (clustered by individual and survey) in parentheses.

future returns[6]. Finding evidence of overreaction is unsurprising for two reasons. First, also using Coibion and Gorodnichenko (2015) regressions, Bordalo et al. (2018) present evidence that overreaction of individual forecasters is prevalent across a wide range of macroeconomic variables. Second, stock returns are very difficult to predict. Campbell and Thompson (2007) show predictive regressions fail to do better than the historical average unless augmented with theoretical restrictions. This points to a weak form of the efficient market hypothesis according to which one cannot use information to which typical U.S. citizens have access to form a forecast more accurate than forecasting the average return would be. In the previous section, we document that expectations react to recent returns and economic news on TV with sign and magnitude varying by group, and that they have sizable unexplained variation survey

[6]Suppose forecasts and returns are uncorrelated and covariance stationary. Then δ equals exactly -0.5:

$$\begin{aligned}\delta &= \frac{cov(FE_t, FR_t)}{var(FR_t)}\\ &= \frac{cov(R_{t\to t+h} - F_t R_{t\to t+h}, F_t R_{t\to t+h} - F_{t-1} R_{t\to t+h})}{var(F_t R_{t\to t+h} - F_{t-1} R_{t\to t+h})}\\ &= -\frac{var(F_t R_{t\to t+h}) - cov(F_t R_{t\to t+h}, F_{t-1} R_{t\to t+h})}{2\cdot var(F_t R_{t\to t+h}) - 2\cdot cov(F_t R_{t\to t+h}, F_{t-1} R_{t\to t+h})} = -\frac{1}{2}\end{aligned}$$

to survey on top of that. The results of Table 6 indicate that this variation in expectations is a form of overreaction.

4 Conclusions

We have analysed an unusually long panel of households' probabilistic stock market expectations collected in the RAND American Life Panel. Our first step was to document a number of key facts in these data, several of which have been known from other datasets and thus help establishing comparability. First, average beliefs are pessimistic relative to historical returns. Second, the dispersion of beliefs is very large, both across individuals in the cross-section and within individuals over time. Third, part of the variation over time is related to the fact that on average, beliefs extrapolate recent trends on the stock market. Fourth, individuals base their expectations for stock returns mostly on the state of the economy and the tone of recent media reports is positively related to average expectations. Fifth, the beliefs of financially sophisticated and knowledgeable individuals are more optimistic. Sixth, a non-trivial fraction of reported beliefs suffers from inconsistencies, part of which may be related to the fact that individuals truly have no quantitatively well-formed expectations. Finally, inconsistent beliefs are found less often for individuals who are financially sophisticated and knowledgeable.

Taking these facts as our point of departure, we have specified a simple model that relates beliefs to past returns and the tone of economic news. We have allowed for heterogeneity by first classifying individuals into one of five groups using the k-means clustering algorithm and then estimating the model separately for each group. The diagnostic test of Dzemski and Okui (2018) revealed that unit-wise confidence sets are small and that in 60% of the cases, they include the group we estimate individuals to be in. Only 12% behave in a way that is not captured by any of our groups, so that their confidence set is empty. This is despite the fact that our approach makes it difficult for the specification test in the sense that it is based on a very different statistic than what is used by the clustering algorithm.

Of our five groups, we have labelled the two polar cases in terms of optimism "pessimists" (annual return expectations well below zero, little reaction of expectations to either returns or news, average values for literacy indices) and "sophisticates" (annual return expectations close to the historical average, small positive reactions to recent returns and news, high scores on literacy / knowledge and few inconsistencies). In between, the “extrapolators” and “mean reverters” expect returns of around zero, have average literacy scores and errors, but they differ sharply in their reaction to returns and news. The extrapolators expect recent trends of both to continue, whereas mean reverters think that the opposite will happen. Finally, the group of “ignorants” stands out from the rest in that they do not seem to be very interested in financial matters, which results in frequent fifty-fifty answers to probabilistic expectations questions. On an ensuing question about whether these answers are supposed to express actual probabilistic judgements or general epistemic uncertainty, they often state the latter. Beliefs and their heterogeneous trajectories are reflected in predicted trading patterns. Our results are robust to different modelling assumptions in a number of directions. None of the five groups passes a rational expectations test; they all overreact in one way or another to recent information.

The evidence that households’ expectations about the development of the stock market are heterogenous is overwhelming; Giglio et al. (2019) is a recent contribution and contains a good overview of previous studies. We have shown that part of this can be traced to heterogenous expectations formations processes. In particular, the much longer time series has allowed us to go beyond the early contribution by Dominitz and Manski (2011) and classify individuals based on a statistical algorithm as opposed to inferring it from two observations only. With a similar structure of data in the Netherlands, Heiss et al. (2019) estimate a finite mixture model of three expectation types based on Dominitz and Manski (2011). Their results are broadly in line with ours. They estimate about one fifth of the population each to be mean reverters and extrapolators (“persistence types” in their terminology), which is very much in line with our estimates. Their random walk type would then encompass sophisticates, pessimists, and ignorants in our analysis in the sense that all these types react very little to changes in stock market returns. This seems plausible given how the two models are specified, but it would be worthy of a more detailed investigation.

These findings have important implications for explaining stock market participation and for asset pricing models. For example, Barberis et al. (2015) develop an asset pricing model with extrapolative investors in addition to rational market participants. The results of Heiss et al. (2019) and of this paper suggest that even more investor types deserve such attention.

References

Adam, Klaus, Albert Marcet, and Johannes Beutel (2017). "Stock price booms and expected capital gains". In: *American Economic Review* 107, pp. 2352–2408.

Adam, Klaus, Albert Marcet, and Juan Pablo Nicolini (Feb. 1, 2016). "Stock Market Volatility and Learning". In: *The Journal of Finance* 71.1, pp. 33–82.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer (Jan. 2015). "X-CAPM: An Extrapolative Capital Asset Pricing Model". In: *Journal of Financial Economics* 115, pp. 1–24.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2017). "Discretizing Unobserved Heterogeneity". IFS Working Paper W17/03.

Bonhomme, Stéphane and Elena Manresa (2015). "Grouped patterns of heterogeneity in panel data". In: *Econometrica* 83.3, pp. 1147–1184.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2018). *Over-reaction in macroeconomic expectations.* Tech. rep. National Bureau of Economic Research.

Bruin, Wändi Bruine de, Baruch Fischhoff, Susan G. Millstein, and Bonnie L. Halpern-Felsher (Jan. 2000). "Verbal and Numerical Expressions of Probability: "It's a Fifty–Fifty Chance"". In: *Organizational Behavior and Human Decision Processes* 81.1, pp. 115–131.

Campbell, John Y and Samuel B Thompson (2007). "Predicting excess stock returns out of sample: Can anything beat the historical average?" In: *The Review of Financial Studies* 21.4, pp. 1509–1531.

Coibion, Olivier and Yuriy Gorodnichenko (Aug. 2015). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". In: *American Economic Review* 105.8, pp. 2644–78.

Dominitz, Jeff and Charles F. Manski (1997). "Using Expectations Data to Study Subjective Income Expectations". In: *Journal of the American Statistical Association* 92.439, pp. 855–867.

— (2007). "Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study". In: *Journal of the European Economic Association* 5.2/3, pp. 369–379.

— (2011). "Measuring and Interpreting Expectations of Equity Returns". In: *Journal of Applied Econometrics* 26.3, pp. 352–370.

Dzemski, Andreas and Ryo Okui (2018). "Confidence Set for Group Membership". Available at `https://adzemski.github.io/research/`.

Enke, Benjamin and Thomas Graeber (2019). "Cognitive Uncertainty". In:

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen P. Utkus (2019). "Five Facts About Beliefs and Portfolios". Available at: `http://dx.doi.org/10.2139/ssrn.3336400`.

Greenwood, Robin and Andrei Shleifer (Mar. 2014). "Expectations of Returns and Expected Returns". In: *Review of Financial Studies* 27.3, pp. 714–746.

Heiss, Florian, Michael Hurd, Maarten van Rooij, Tobias Rossmann, and Joachim K Winter (2019). “Dynamics and heterogeneity of subjective stock market expectations”. In:

Hudomiet, Péter, Michael D. Hurd, and Susann Rohwedder (2018). “Measuring Probability Numeracy”. RAND Labor and Population Working Paper, WR-1270.

Hurd, Michael D. (Sept. 2009). “Subjective Probabilities in Household Surveys”. In: *Annual Review of Economics* 1.1, pp. 543–562.

Hurd, Michael D. and Susann Rohwedder (Oct. 2011). “Stock Price Expectations and Stock Trading”. RAND Labor and Population Working Paper 938.

Hurd, Michael D., Maarten C. J. van Rooij, and Joachim Winter (2011). “Stock Market Expectations of Dutch Households”. In: *Journal of Applied Econometrics* 26.3, pp. 416–436.

Kezdi, Gabor and Robert J. Willis (2008). “Stock market expectations and portfolio choice of American households”.

Kleinjans, Kristin J. and Arthur Van Soest (2014). “Rounding, Focal Point Answers and Nonresponse to Subjective Probability Questions”. In: *Journal of Applied Econometrics* 29.4, pp. 567–585.

Lusardi, Annamaria and Olivia S. Mitchell (2007). “Financial Literacy and Retirement Planning: New Evidence from the RAND American Life Panel”. University of Pennsylvania, The Wharton School, Pension Research Council Working Paper WP2007-33.

— (Mar. 2014). “The Economic Importance of Financial Literacy: Theory and Evidence”. In: *Journal of Economic Literature* 52.1, pp. 5–44.

Malmendier, Ulrike and Stefan Nagel (2011). “Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?” In: *Quarterly Journal of Economics* 126.1, pp. 373–416.

Manski, Charles F. (2004). “Measuring Expectations”. In: *Econometrica* 72.5, pp. 1329–1376.

Manski, Charles F. and Francesca Molinari (2010). “Rounding probabilistic expectations in surveys”. In: *Journal of Business and Economic Statistics* 28, pp. 219–231.

Pedregosa, F. et al. (2011). “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.