ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

von Gaudecker, Hans-Martin; Wogrolly, Axel

Working Paper

The dynamics of households' stock market beliefs

CESifo Working Paper, No. 7602

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: von Gaudecker, Hans-Martin; Wogrolly, Axel (2019) : The dynamics of households' stock market beliefs, CESifo Working Paper, No. 7602, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/198962

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo WORKING PAPERS

7602
2019

April 2019

The dynamics of households' stock market beliefs

Hans-Martin von Gaudecker, Axel Wogrolly

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

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

CESifo *Working Paper No. 7602*
Category 12: Empirical and Theoretical Methods

The dynamics of households' stock market beliefs

Abstract

We analyse a long panel of households' stock market beliefs to gain insights into the nature of their expectations formation processes. We classify respondents into one of five groups based on their data and estimate group-wise models of expectations formation. Two of the groups are at opposite extremes in terms of optimism: Pessimists who expect substantially negative returns and financially sophisticated individuals whose expectations are close to the historical average. Two groups expect 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 poor probability numeracy; its individuals are not willing or able to quantify their beliefs about future returns. None of the estimated belief formation processes passes a rational expectations test.

JEL-Codes: D830, D840, D140, C380.

Keywords: stock market expectations, household finance, heterogeneity, clustering methods.

Hans-Martin von Gaudecker
University of Bonn
Germany – 53012 Bonn
hmgaudecker@uni-bonn.de

Axel Wogrolly
University of Bonn
Germany – 53012 Bonn
axelwogrolly@uni-bonn.de

April 3, 2019
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/CRCTR 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 DFG under CRC-TR 224.

1 Introduction

Understanding households' stock market expectations is critical for models of portfolio choice, life-cycle behaviour, 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 heterogeneous belief formation processes. 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.

To analyse heterogeneity in belief formation, we employ the strategy proposed in Bonhomme, Lamadon, and Manresa (2017). In a first step, we employ the k-means clustering algorithm to assign individuals to groups based on their data. This is very similar in spirit to Dominitz and Manski (2011), who classify individuals into one of three pre-defined types based on two time series observation in the Michigan Survey of Consumers. In light of the fact that stated beliefs are often imprecise (Gouret and Hollard, 2011; Drerup, Enke, and von Gaudecker, 2017), this may lead to classifying some individuals based on noise rather than swings in their beliefs. Thanks to a much longer time series dimension—26 observations on average—we are able to use a statistical classifying procedure and consider five groups in our main specification. Using less groups mixes individuals with very different economic behaviours; adding more leads to little additional insights. We find that these groups are rather stable when varying important features of the sample or of the classifying procedure. 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 16% 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 who stand out for having better knowledge of financial markets and the stock market in particular. Their expectations are close to the historical performance of the stock market and respond slightly positively to recent returns and news about the economy. At the other extreme of average expectations, we estimate one group with substantially negative return expectations, little reaction to returns or news, and average values both for literacy indices and 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; the other expects them to revert again. The last group stands out from the rest in that its members do not care much about the behaviour of the stock market, their belief measures often violate the laws of probability calculus, and when given the chance, they frequently state that their answers are better described as reflecting epistemic uncertainty as opposed to subjective probability judgements. In a final step, we use the method of Coibion and Gorodnichenko (2012) to test

whether the expectations of any of our groups could be characterised as rational. We find that this is not the case; all overreact to current information.

Our findings are consistent with recent evidence from a mixed survey-administrative dataset in Giglio et al. (2019), who document persistent heterogeneity in the levels of beliefs, which is difficult to explain with observable characteristics. However, we also find that for the two groups in the middle of the optimism/pessimism distribution, both data averages and predictions frequently cross over the period of our analysis. This is driven by opposing responses to information, with one group extrapolating and the other mean-reverting, around a similar expected return.

The rest of the paper is organised as follows. Section 2 describes our data, connects it to prior literature and establishes the key stylised facts for our data. In section 3, we outline our empirical strategy and 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 and stylised facts

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 A.1 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

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).

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.

In terms of individual characteristics beyond standard socio-economic variables our data includes 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 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 contains 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 A.1 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 $[0.2, \infty]$. We limit the boundaries of the open intervals to the 1st and $99^t h$ percentiles of the historical distribution of the Dow Jones' return (-0.32 and 0.43, respectively) and set all conditional means to the interval midpoints. 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, 40% of responses in the data of Hurd, Rooij, and Winter (2011)—and we devote Section 2.7 entirely to their analysing them.

Previous authors have only used the probability of a positive return even though more measures would have been available (Dominitz and Manski, 2007). We will do the same in the form of a robustness check.

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

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

Units in percentage points.

variation across the different points of the distribution function appears reasonable and all measures exhibit substantial variation across individuals.

2.2 On average, beliefs are pessimistic compared to historical returns

A comparison of the average subjective beliefs with the distribution of historical returns reveals that the individuals in our sample are pessimistic about the stock market. This finding is in line with Hurd's 2009 summary of various studies and data as well as Hurd, Rooij, and Winter's 2011's report for Dutch households.

In Table 3 we collected expected returns and probabilities for returns exceeding -20%, 0% and 20% from the historical data and compare them with the average subjective beliefs. Individuals are too pessimistic by 23 and 28 percentage points respectively that the Dow Jones will not collapse and that it will increase. The fact that individuals seem to be too optimistic that the Dow Jones will increase by 20 percent or more relative to empirical frequencies should probably not be taken at face value. If we drop individuals who exhibit monotonicity violations from the sample, the difference changes sign in line with the other values. In sum, relative to the historical distribution, individuals are, on average, too pessimistic.

Table 3: Historical returns vs. beliefs about returns

	Historical Averages	Subjective Beliefs	Difference
$E[R_{t\to t+12}]$	7.3	0.5	6.9
$\Pr(R_{t\to t+12}) > -0.2$	97.1	74.6	22.5
$\Pr(R_{t\to t+12}) > 0$	72.1	44.0	28.1
$\Pr(R_{t\to t+12}) > 0.2$	23.5	26.8	-3.3

Units in percentage points. The historical averages $\Pr(R_{t\to t+12} > x)$ are estimated using the empirical frequency $T^{-1}\sum_t^T 1\{R_{t\to t+12} > x\}$ for yearly returns of the Dow Jones between 1950 and 2016. Beliefs are within-person means.

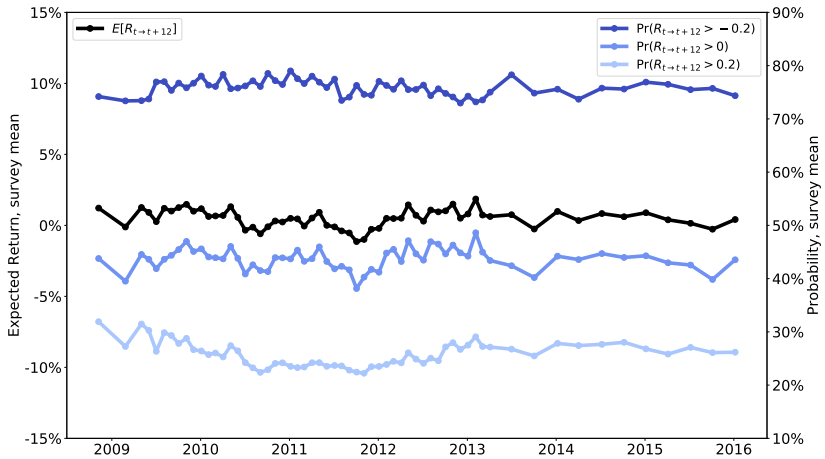
2.3 Beliefs exhibit significant dispersion within and across individuals

Table 2 has already shown the substantial variation in average beliefs across individuals. The same holds true for the variation within persons across time with comparable magnitudes (see Table A.2).

One notable feature is that for both within and between-subject differences, the variation is largest for the first and arguably most intuitive question, i.e., $\Pr(R_{t\to t+12} > 0)$. In Section 2.6, we confirm that individual characteristics have most predictive power for variation in this measure of an individual's beliefs about the future of the Dow Jones.

Figure 1 shows that the substantial belief variation over time we find at the individual level largely cancels out if beliefs are averaged across subjects. Unless the within-variation is unsystematic, this is an indication that average beliefs averages mask substantial heterogeneity in belief dynamics.

Figure 1: Average beliefs over time



Depicted series are within-survey means. The left y-axis displays the scale for the expected returns, the right y-axis displays the scale for the three probabilities.

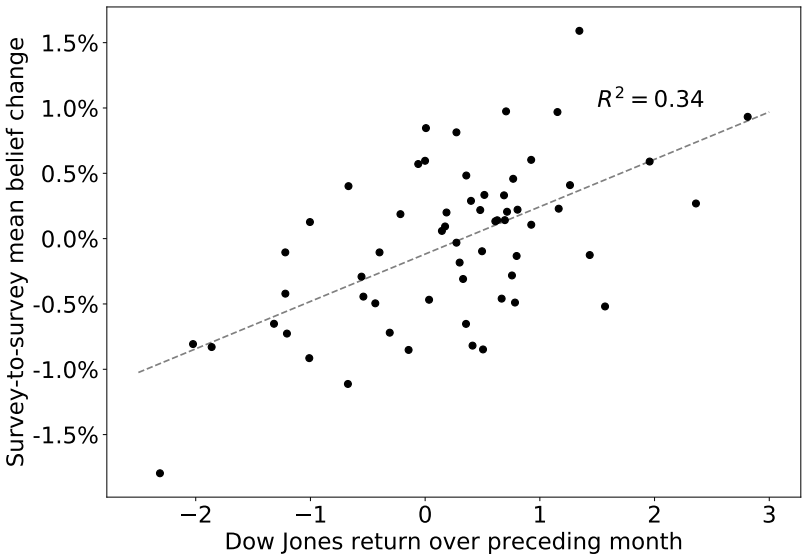
2.4 On average, beliefs extrapolate recent stock market returns

A recent literature has documented that average return expectations covary with recent stock market movevements. 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".

In order to investigate the extent to which the same finding can be replicated in our data, we compute the first differences of expected returns, average across individuals for every survey wave and plot them against the Dow Jones' return over the past month in Figure 2. We checked a variety of other periods and one or two months produced the highest correlations. The results show that on average, beliefs indeed extrapolate recent changes into the future. An increase in the Dow Jones' returns over the past month of 4.3 percentage points (one standard deviation of the monthly return over the period on which our regression is based) is associated with a 0.36 percentage point higher expectation on the return over the next year. For comparison, 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.

Figure 2: Average beliefs extrapolate stock prices

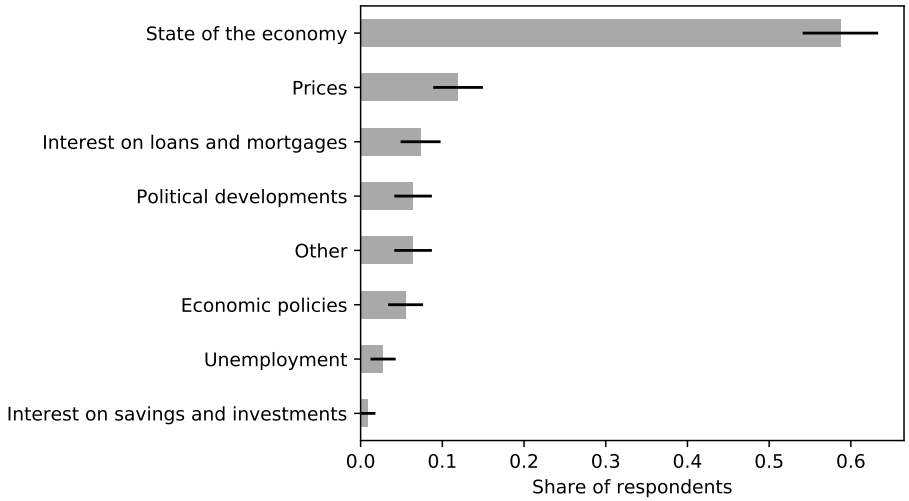


Between-subject mean belief changes (in expected returns) are plotted against the standardised return of the Dow Jones.

2.5 On average, beliefs follow the tone of recent media reports

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

Figure 3: What respondents think most about when contemplating future stock prices



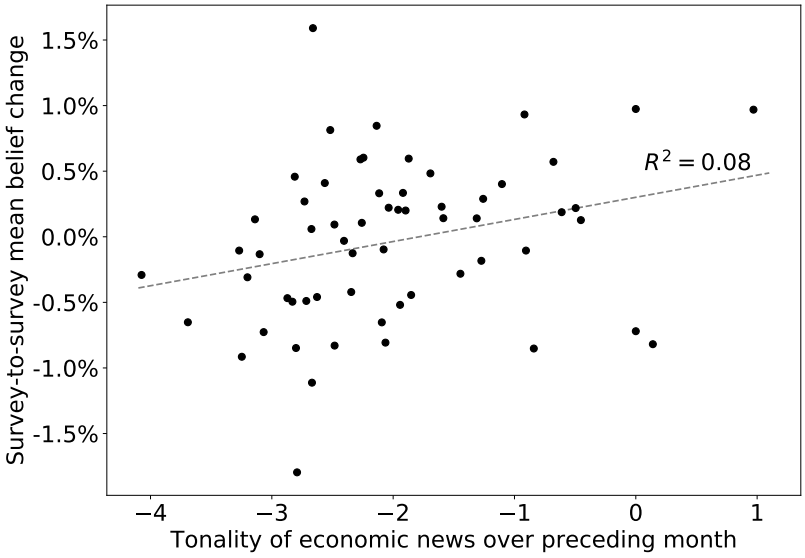
Only includes individuals who overlap with our main sample, $N = 114$.

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 the beginning of this Section 2) lead us to include additional information that subjects may use to infer the state of the economy. 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

[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?"

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: $N_{t-1\to t} := \frac{\sum_{d\in[t-1,t]} 1\cdot\text{pos}_d + 0\cdot\text{neu}_d - 1\cdot\text{neg}_d}{\sum_{d\in[t-1,t]} \text{pos}_d + \text{neu}_d + \text{neg}_d}$.

Figure 4: On average, beliefs follow the tone of recent media reports



Between-subject mean belief changes (in expected returns) are plotted against the standardised measure of economic news tonality.

Figure 4 shows a scatter plot of between-subject mean belief changes in expected returns (survey to survey) against this measure. As with past returns of the Dow Jones in the previous section, we find a positive relation. The fact that news cannot account for as much variability in average belief changes as returns should not be surprising as one would expect the coders of Media Tenor to be a noisier aggregator of information than the stock market. This is exacerbated by smallish sample sizes in some months.

2.6 Beliefs of financially sophisticated and knowledgeable individuals are more optimistic

To get a sense of what is driving persistent level differences in beliefs, we once more average beliefs within individuals and then regress them on individual-level characteristics. Table 6 reports the results.

Table 4: Predictors of average beliefs

	$E[R_{t\to t+12}]$	$\Pr(R_{t\to t+12} > -0.2)$	$\Pr(R_{t\to t+12} > 0)$	$\Pr(R_{t\to t+12} > 0.2)$
Follows stock market	1.16**	0.29	3.56**	1.61
	(0.53)	(1.23)	(1.61)	(1.30)
Understands stock market	0.41	0.02	2.12	0.00
	(0.51)	(1.13)	(1.50)	(1.22)
Knowledge of past returns: Don't know	1.45**	1.31	4.30**	1.41
	(0.61)	(1.56)	(1.88)	(1.61)
Knowledge of past returns: Magnitude too large	4.84***	2.01	13.35***	8.12***
	(1.07)	(2.15)	(2.70)	(2.73)
Knowledge of past returns: Sign and Magnitude correct	3.06***	3.26**	9.23***	2.71*
	(0.62)	(1.47)	(1.87)	(1.54)
Probability Numeracy	1.09***	0.22	4.07***	0.99
	(0.26)	(0.56)	(0.78)	(0.67)
Financial Knowledge	0.34	0.71	2.09***	-1.12
	(0.25)	(0.64)	(0.81)	(0.71)
Financial Numeracy	0.12	0.37	1.19	-0.94
	(0.25)	(0.57)	(0.76)	(0.67)
R^2	0.22	0.11	0.32	0.05
N	805	805	805	805

OLS estimates. Standard errors (robust) in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. Omitted categories are 'Does not follow stock market', 'Does not understand stock market', and 'Knowledge of past return: Wrong sign given'. Dependent variables are within-person means in percentage points. Measures of Financial and probability numeracy are standardised. Regressions include the following controls: Gender, age, age squared, education, ethnicity, family income, whether respondents have ever held stock in their liquid portfolio. The full set of results can be found in the Online Appendix, Table A.3.

We focus on variables that capture the extent to which people are involved with, and have knowledge of, the stock market and financial matters more generally. All regressions included controls. The signs of the significant predictors confirm what we would expect: A better knowledge of past returns and financial matters, as well as following the stock market, are associated with more optimistic beliefs. Meanwhile, self-assessed understanding does help much to predict beliefs conditional on knowledge of past returns and financial knowledge.

Knowledge of past returns, our most direct measure of an individuals' information set, is the strongest predictor for expected returns and for all three probabilistic beliefs. Relative to respondents who state the wrong sign for a Dow Jones return over the past year, individuals who give the correct sign and magnitude (or overestimate the latter) are 9-13 percentage points more optimistic that the Dow will increase over the coming year; they expect returns that are 3-5 percentage points higher on average.

Higher probability numeracy and financial knowledge also predict optimism in the belief that the Dow will increase. A one standard deviation increase in these scores predicts increases in the beliefs that the Dow will rise of 4 and 2 percentage points. That probability numeracy is associated with belief levels conditional on various indicators measuring what people know about the stock market points at measurement error in stated beliefs.

As noted before, the predictive power of the covariates is much higher for the probability of a positive return than for the other two points on the distribution function; the R^2 differs by factors of three to five. We take this as additional evidence pointing towards higher noise levels for the events of the Dow Jones rising or falling by at least 20%. Put differently, we should not take all stated beliefs at face value.

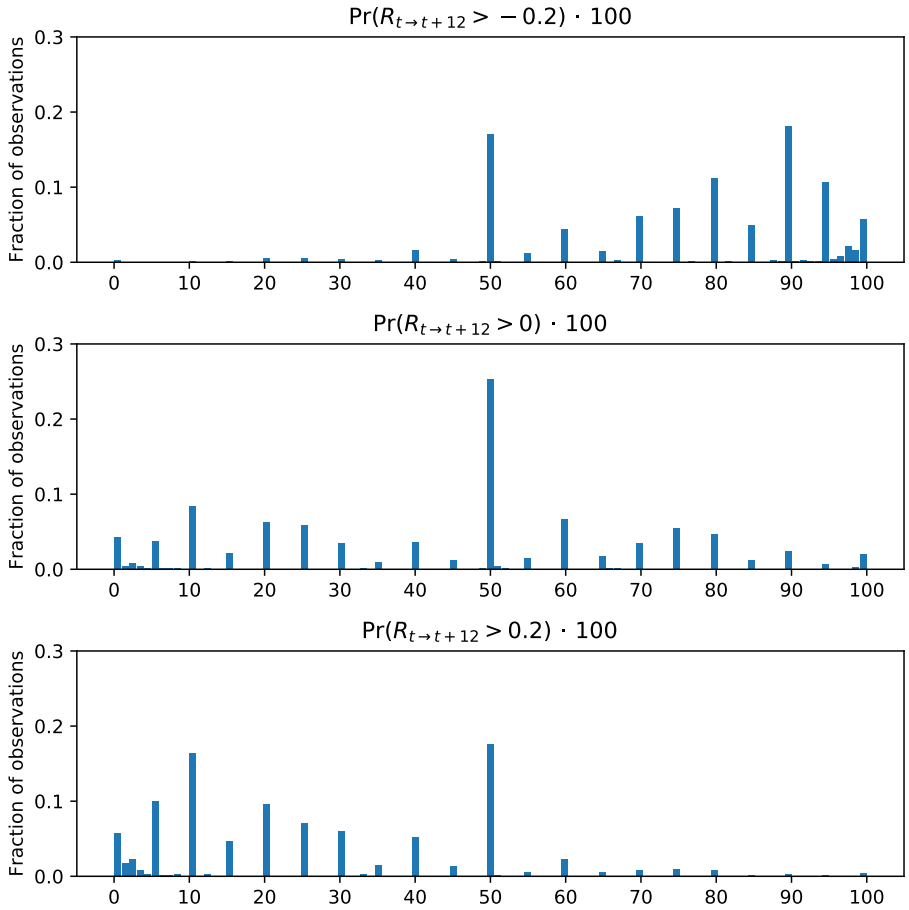
2.7 Stated beliefs vary in their information value

Measurement error and / or imprecision in stated beliefs have concerned researchers for a long time. Two particularly prevalent phenomena are rounding of stated probabilities and the previously-mentioned monotonicity violations. We regard both as indications that stated measures are less informative about what an individual thinks about the stock market, similar in spirit to Drerup, Enke, and von Gaudecker (2017).

Figure 5 shows histograms of the beliefs with 1-percent bins. Most beliefs are rounded to the nearest multiple of 5% or 10%, and that answers equalling 50% are particularly frequent. The middle Panel of Figure 5 looks very similar to Figure 3 in Hurd, Rooij, and Winter (2011). These basic facts on rounding have been documented for a long time, Manski and Molinari (2010) and Kleinjans and Soest (2014) are recent contributions and modelling suggestions. Rounding suggests individuals are either not willing to exert the effort to express a precise belief, or that their beliefs themselves are imprecise.

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 confronts 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 are simply unsure. 47% of responses to this question indicated that they are unsure, not that they judge the probabilities to be equal. As one would expect if people do not have a well formed belief, the stated probability for $\Pr(R_{t\to t+12} \geq 0.2)$ and $\Pr(R_{t\to t+12} \leq -0.2)$ also equalled 50% about half of the time in in that case. By contrast, for the 53% of responses indicating that a probability of 50% means they find an increase and a decrease equally likely, the other two probabilities equalled 50% only one third of the time.

Figure 5: Distributions of belief variables



The figures depict histograms of belief variables with 1-percent bins. Data is pooled across surveys, including only individuals with at least five sets of belief measurements.

A striking irregularity in measured beliefs are monotonicity violations. Similarly to rounding, this is in line with what previous studies of probabilistic expectations have found (e.g. Hurd, Rooij, and Winter, 2011; Hudomiet, Kezdi, and Willis, 2010). Our raw beliefs data consists of 3 points on the cumulative distribution function: -0.2, 0 and 0.2. There was no reminder that stated beliefs have to (weakly) increase along these points, and hence answers can violate the monotonicity property of the cumulative distribution function. Stated beliefs that are not monotone are incoherent, and thus cannot be regarded as very informative about what people believe will happen with the Dow Jones. To a somewhat lesser extent, this is true for weakly but not strongly monotone beliefs as well. While compatible with probability calculus, such answers suggest respondents think there is no chance the return of the Dow could be

between -20% and 0% or 0% and 20%, even as they do think there is a chance returns could be smaller or larger than that. Table 5 shows the incidence of monotonicity violations in our data. Around 70% of stated beliefs sets are strictly monotone between the points -0.2 and 0 as well as 0 and 0.2, making for 57% that satisfy both checks.

Table 5: Prevalence of monotonicity violations

	From -0.2 to 0	From 0 to 0.2	Either
Not monotone	0.08	0.07	0.15
Weakly but not strictly monotone	0.18	0.23	0.28
Strictly monotone	0.74	0.70	0.57

Table shows fraction of beliefs satisfying each listed monotonicity status.

Table 5 shows that a substantial number of people give answers that do not obey the rules of probability calculus or seem implausible. The propensity to give monotonicity violating answers may be thought of as being determined by the effort give when answering the survey and by how much effort is required to avoid errors and give reasonable answers. While we cannot observe effort, people familiar with financial markets, in particular stock markets, should find it easier to avoid mistakes. In addition, such people are more likely to hold precise beliefs in the first place, as their information set is richer. Knowledge of probability calculus and familiarity with using probabilities to indicate uncertainty can also be expected to reduce the incidence of nonsensical answers. Both of these are likely positively related to effort, as people are more willing to do tasks they are good at and interested in.

2.8 Beliefs of financially sophisticated and knowledgeable individuals are more consistent

To investigate what drives monotonicity violations, epistemic uncertainty, and rounding we use measures of probability numeracy, financial numeracy and engagement with the stock market along with typical characteristics such as gender, age, education, income and ethnicity. As before, we collapse the time dimension of our data. We compute an individual's average propensity to express non-monotone or weakly monotone beliefs, their average propensity to say that their 50% beliefs mean they are unsure as opposed to a subjective probability (if individuals did not see this follow up question because they did not give a 50% answer, we

assume their answer is a subjective probability) and their average propensity to give answers that are multiples of 5% as dependent variables. We regress these on personal characteristics. Kezdi and Willis (2008) and Gouret and Hollard (2011) find no relationship between the propensity to give problematic answers and general personal characteristics, but we find strong relationships between financial and probability numeracy and non-monotone or epistemically uncertain beliefs.

Table 6: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 5%
Follows stock market	-0.05**	-0.02	0.02
	(0.02)	(0.01)	(0.01)
Understands stock market	-0.03	-0.02**	0.01
	(0.02)	(0.01)	(0.01)
Knowledge of past returns: Don't know	-0.07**	0.01	-0.02
	(0.03)	(0.02)	(0.02)
Knowledge of past returns: Magnitude too large	-0.05	0.01	-0.02
	(0.04)	(0.03)	(0.03)
Knowledge of past returns: Sign and Magnitude correct	-0.08***	-0.01	-0.02
	(0.03)	(0.02)	(0.02)
Probability Numeracy	-0.06***	-0.01	0.00
	(0.01)	(0.01)	(0.01)
Financial Knowledge	-0.06***	-0.03***	-0.01
	(0.01)	(0.01)	(0.01)
Financial Numeracy	-0.04***	0.00	0.01
	(0.01)	(0.01)	(0.01)
R^2	0.31	0.13	0.25
N	805	805	805

OLS estimates. Standard errors (robust) in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. Omitted categories are 'Does not follow stock market', 'Does not understand stock market', and 'Knowledge of past return: Wrong sign given'. Measures of financial and probability numeracy are standardised. Regressions include the following controls: Gender, age, age squared, education, ethnicity, family income, whether a respondent has ever held stock in a liquid portfolio, whether respondents found the survey interesting, the duration it took a respondent to answer, how many times they participated. For the dependent variable non-monotone, we additionally include the average belief level that the Dow Jones will decrease.

In line with the earlier discussion, the regression results in Table 6 demonstrate that following the stock market, having accurate knowledge of historical returns, probability numeracy, and financial numeracy all predict that an individual is less likely to express beliefs afflicted by monotonicity errors. The most important predictors for individuals to state that their expressed beliefs indicate likelihoods are self-assessed understanding of the stock market, probability numeracy and financial numeracy. One interpretation of these associations is that richer information sets and greater understanding lead to more precise beliefs, and lower the costs of stating beliefs in the survey, which reduces the incidence of errors. Greater familiarity with probabilities also lowers errors and makes it more likely that individuals use 50% answers to

indicate equal likelihoods. Rounding, measured as the fraction of answers that are multiples of 5, is not systematically predictable with our indicators of sophistication. This is one reason why we follow Hurd, Rooij, and Winter (2011) and do not adjust for rounding as in Manski and Molinari (2010).

3 Modelling and estimating heterogeneous belief formation processes

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, these facts point towards putting between-person heterogeneity at the centre of a model of belief formation. 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 how they interpret new information, and 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.

We first outline our empirical strategy and a diagnostic test for the classification of individuals to groups. We then describe our main results and 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 Empirical strategy

Our main specification for belief formation is a linear model of the form:

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

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 (see Subsection 2.5).

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 scikit-learn Python library (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 our main specification, we choose the following parameter G, classification variables, and sample restrictions. We use five groups because this was the minimum number of groups where no economically meaningful intergroup differences were blurred; larger numbers led to apparent overfitting and little additional insights. We will be more precise on this below in Sections 3.3 and particularly in 3.4, where we also consider alternative choices for G.

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. This makes for a total of twelve time-constant moments that vary across individuals. We make this choice because it just uses raw data (as opposed to, for example, using the expected returns, which entail a number of assumptions as detailed in Section 2.1) and because these are the key moments that should be informative on group-level heterogeneity, as required for the analysis in Bonhomme, Lamadon, and Manresa (2017).[2] Section 3.4 reports on results when adding measures of belief consistency to the classification step.

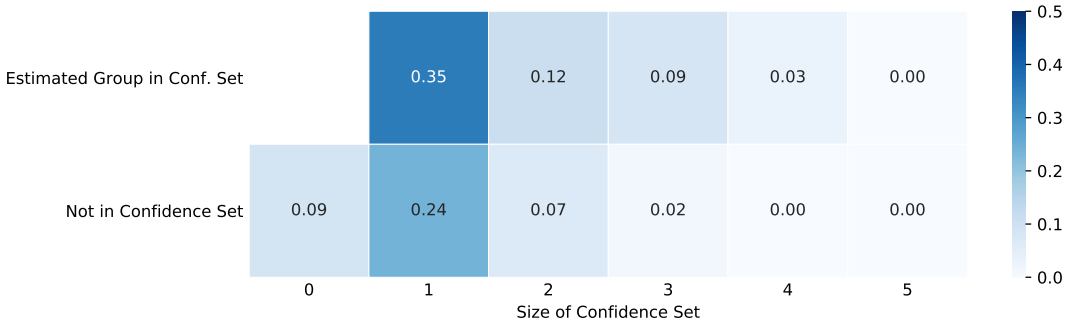
We discard individuals for whom we have less than five time-series observations in order to have a certain amount of faith in the moments used for classification. In the Online Appendix, we show that our results are similar when using only two observations per individual—the minimum number allowing the calculation of our preferred moments—or strengthening the requirement to fifteen observations, which would guarantee that Equation (1) is identified at the individual level for the specification with $L = 6$.

3.2 Diagnostics

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 g_i^0 is g for all groups $1, \ldots, G$. The elements of the confidence sets are then those groups for

[2] 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.

Figure 6: Unit-wise 10% confidence sets by size and inclusion of estimated group



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

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).

Figure 6 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).

For nine percent of our sample, the confidence sets are empty. This can occur if none of the groups provides a good fit. To see this, we examine the statistic Dzemski and Okui (2018) use to test the null hypothesis that $g_i^0 = g$. Rejection means that g will not be in the confidence set. The infeasible statistic uses the difference in means-adjusted squared residuals, which can be written as $d_{i,t}(g,h) = -u_{i,t}x_{i,t}^T(\theta_g - \theta_h) + \left(\theta_g - \theta_{g^0}\right)^T x_{it}x_{it}^T(\theta_g - \theta_h)$. For a given group g, $\frac{1}{T_i}\sum_t E[d_{i,t}(g,h)]$ is estimated for all h in $\{1,\ldots,G\}\backslash\{g\}$. If at least one element of the resulting vector differs sufficiently from zero, then g will not be in the confidence set. This can detect incorrect group assignments because $E[u_{i,t}|x_{i,t}] = 0$ ensures that the first term of $d_{i,t}(g,h)$ is zero in expectation but the second term is only zero if $\theta_g = \theta_{g^0}$, i.e., if g is the true group. This proedure is repeated for all g and those for which the null hypothesis is not rejected are added to the confidence set. If the number of groups is too low or the model is misspecified for some individuals, the true parameter vector θ_{g^0} will not be among any of the estimated coefficient vectors, $|\theta_g - \theta_{g^0}|$ will be positive for all groups $g \in \{1,\ldots,G\}$ and the test statistic will reject in all cases. In our case, lowering G will indeed increase this number ($G = 4$ leads to empty confidence sets for 11% of the sample, see Figure C.16 in the Online Appendix, $G = 2$ to 23%). Incidentally, however, the number remains the same when moving to seven groups (Figure C.20, it even slightly increases from 0.088 to 0.094), which leads us to conclude that some individuals' behaviour is either very erratic or difficult to capture with our model. In any case, increasing the number of groups at around our preferred level does not seem to help capturing additional patterns of behaviour but further divides existing ones.

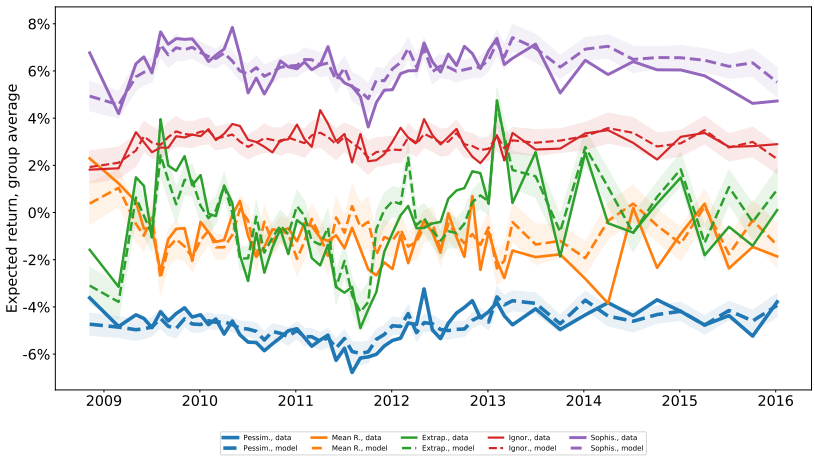
3.3 Results on belief formation processes

We order the 5 groups that emerge from the k-means algorithm by their average expected returns and refer to them as pessimists, mean reverters, extrapolators, ignorants, and sophisticates, respectively. These labels capture the key pattern of behaviour unique to each group. The key results are summarised in three figures. Figure 7 plots the data averages (solid lines) versus the model predictions (dashed lines) of expected returns over time. Figure 8 plots the

reaction of groups to changes in past returns and news, respectively. Finally, Figure 9 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 7 are strikingly different and—except for the mean reverters and extrapolators—hardly ever cross. The reactions to both stocks and news depicted in Figure 8 are substantial and very different. Figure 9 demonstrates that several individual characteristics—including many of those not used for group assignment—vary strongly with the type of group. It is important to note that all group sizes are substantial. The largest group's share is 26% and that of the smallest is 16%.

Figure 7: Data vs. predicted expected return of the Dow Jones index, by group

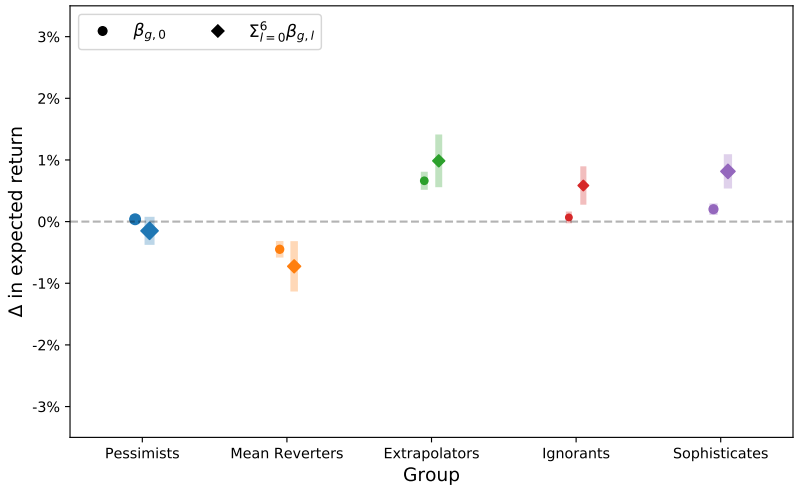


The solid and dashed lines are within survey and group means of individual data points and model predictions. Shaded regions are within survey and group means of individual 95% confidence intervals for the estimated regression function. Line widths are proportional to group sizes.

Pessimists (26% of individuals) consistently expect the return of the Dow Jones to be negative and substantially so. They assess the chances that the Dow Jones will increase to be less than 30% on average. They do not respond to returns and react slightly positively to news. Along the dimensions of knowledge, errors, experience, and sophistication, the pessimists appear to be in the middle of the distribution along with mean reverters and extrapolators. The two apparent exceptions to this—the beliefs of pessimists are less likely to feature monotonicity violations and they are more likely to express subjective probabilities—are probably due

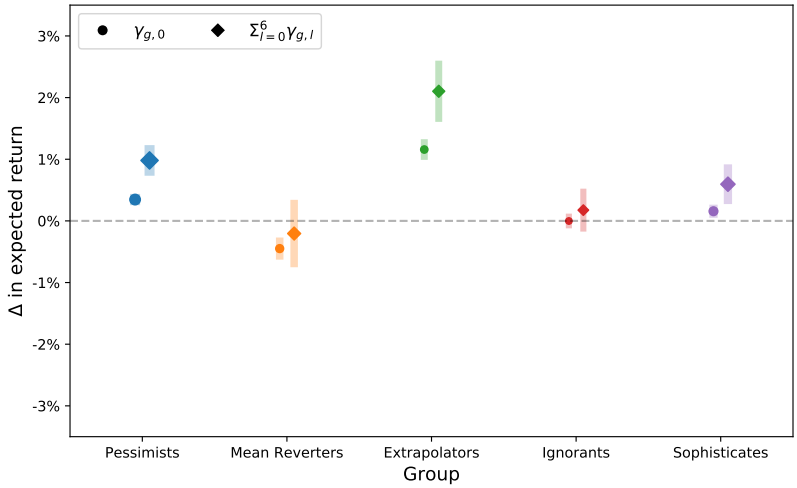
Figure 8: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



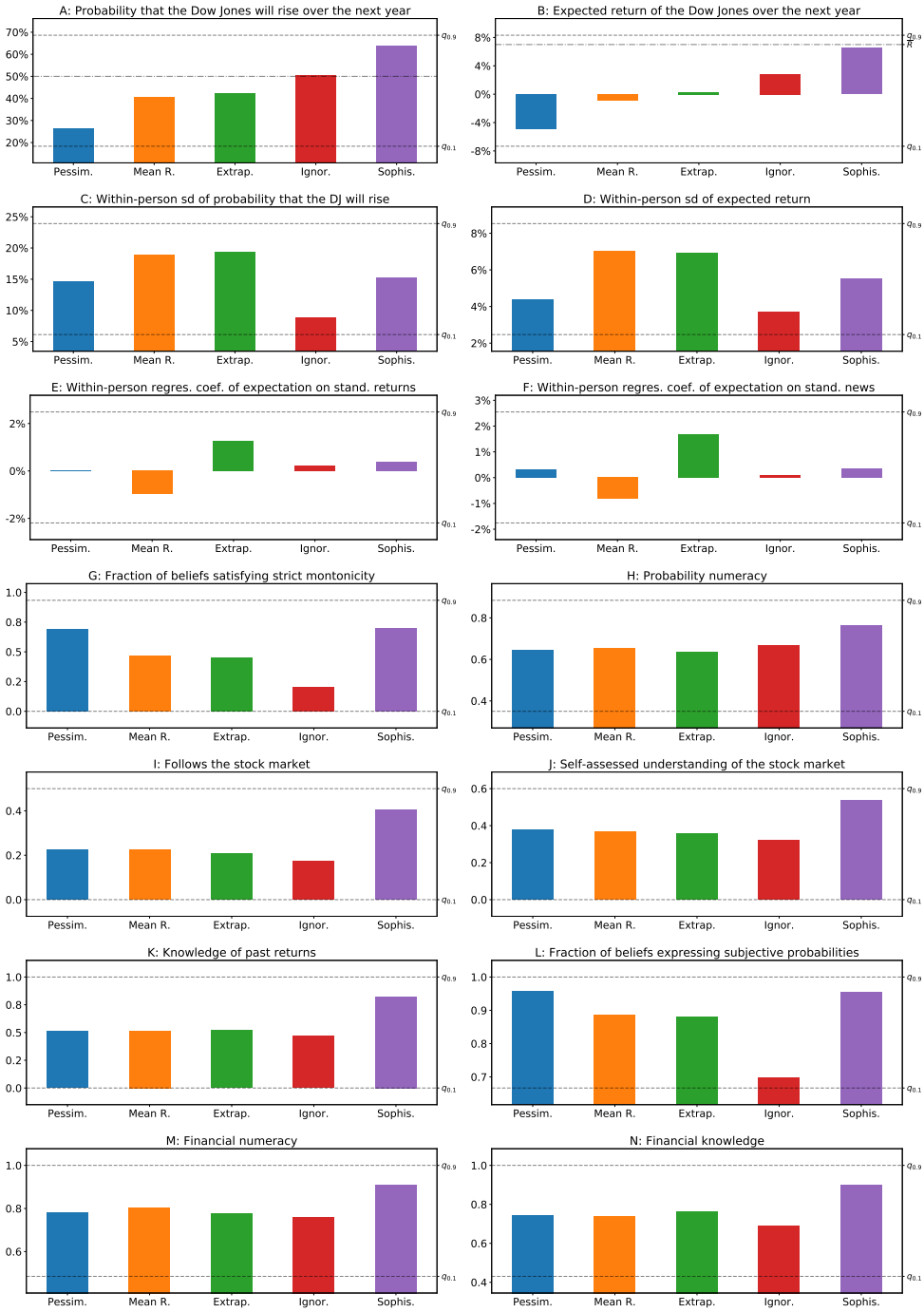
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker and line widths proportional to group sizes.

(b) Effect of past tonality of economic news



Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds, smaller for some panels depending on the avaailability of covariates, see Table 1 depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker and line widths proportional to group sizes.

Figure 9: Characteristics by unobserved heterogeneity group



N = 3030, smaller for some panels depending on the avaailability of covariates, see Table 1. *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).

to somewhat mechanical effects. 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. As 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 $\Pr(R_{t\to t+12} < -0.2) \leq Pr(R_{t\to t+12} < 0)$.

Mean Reverters and Extrapolators (19% and 18% of individuals, respectively) are also rather pessimistic. They expect a return of about 0% and think that the Dow Jones only has a 42% chance to rise. Individuals of these two groups are similar in all characteristics depicted in Figure 9 with the exception of how their beliefs respond to recent returns and economic news. Of course, this is reflected in Figure 8: 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. Hence, the lines in Figure 7 frequently cross and show completely different patterns.

The fact that mean reverters and extrapolators are almost identical 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 C.1 of the Online Appendix), would hide this important dimension of behaviour.

Ignorants (16% 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 2.7%. Compared to the other groups, ignorants are notable for their very low belief variability. Panel C of Figure 9 shows that their average is near the tenth overall percentile. Their beliefs barely covary with returns and news and are most afflicted by monotonicity errors. This group captures individuals who predominantly give 50% answers and often state in the follow-up question that this answer is supposed to express deep uncertainty (Panel L of Figure 9). In line with this interpretation, 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 have low information value, and need to be interpreted with caution.

Sophisticates (22% of individuals), the most optimistic group, expect the Dow Jones to increase over the next year with a probability of 64% and to yield an average return of 6.4%. These values are only slightly lower than the historical performance of the Dow (see Table 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, and they use beliefs to express subjective probabilities most often.

3.4 Discussion and robustness of results

Our preferred model explains more than a quarter of the variation in expected returns. This differs by a factor of one hundred from the same model without unobserved heterogeneity ($R^2 = 0.0023$). Similarly, a model with lots of observed heterogeneity can explain only 12% (see Table C.1). 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 C.2 of the Online Appendix, we relax the requirement of observing at least five sets of belief measures per individual to a minimum of two. The broad pattern of groups remains the same and we can essentially leave the labels in place. There is one exception: The k-means algorithm does not cleanly separate mean-reverters and extrapolators anymore. Instead, it pools half of the individuals classified as extrapolators in our preferred specification with mean-reverters (see Table C.3), creating a group that hardly reacts to returns or news on average. Only twenty percent are left in the group of extrapolators, which now only makes

up 6% of the sample and strongly extrapolates returns. However, standard errors are wide and the group is so small that many periods feature less than fifteen observations, so that we refrain from drawing the respective data averages in Figure C.2. We take this result as cautioning against classifying individuals into types when only few observations are available and the data is noisy.

In contrast to this, 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 C.3 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 C.5), 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 C.4, 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 pattern to Figure 8, if anything, lags seem to have slightly stronger effects towards building up momentum.

The main effect of adding the fraction of individuals' belief sets satisfying strict monotonicity and the fraction of an individuals' beliefs expressing subjective probabilities to the set of variables used for classification is that some groups change size. Section C.5 contains the results. The group of ignorants shrinks by three percentage points and sophisticates gain by four; there is not much action for the other three groups. As the mean belief now gets less relative weight in the classification procedure, it should not be surprising that the groups move somewhat closer together when considering the time series of beliefs in Figure C.13. Note that the groups of ignorants and extrapolators become smaller than fifteen in some survey waves where the sample size was very small (see Table A.1), we leave the respective plots empty.

Also unsurprisingly, the ignorants are now even sharper distinguished in terms of their low scores on monotonicity, their beliefs expressing probability judgements, but also most other variables measuring financial sophistication and knowledge. As in the previously-mentioned alternative specifications, group assignment remains very stable with about 85% of individuals on the diagonal of Table C.8.

Sections C.6, C.7, and C.8 show the results for 4, 7, and 15 groups, respectively. The results further motivate our choice of G. In the case with four groups, the first four groups remain very stable (each retains more at least 95% of its previous members), but the group of sophisticates is distributed across the other groups. The shares are distributed roughly equally across ignorants, pessimists, and the combination of mean-reverters and extrapolators, respectively (see Table C.10). 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 and extrapolators. The other three groups retain more than 85% of their members and all their characteristics remain very stable. Some of the clusters mostly formed by splitting up the extraplators and mean reverters become fairly small and relative to the other lines in Figure C.21, 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 indivduals 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 21 individuals. Note that the diagnostic test from Section 3.2 becomes computationally infeasible.

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), 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 (2012) who specify the information set I_t to be the forecast revision. Let $F_t R_{t\to t+h}$ be the forecast of the return $R_{t\to t+h}$ at time t. Forecast errors are then defined as $FE_t := R_{t\to t+h} - F_t R_{t\to t+h}$ and forecast revisions as $FR_t := F_t R_{t\to t+h} - F_{t-1} R_{t\to t+h}$. Regressing forecast errors on forecast revisions then tests the rationality of expectations. The intercept coefficient measures systematic bias in the level of the forecast, and the sign of the slope coefficient measures whether expectations overreact or underreact to information. If expectations are rational both the intercept and slope coefficient are 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 obtain forecast revisions, 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?". 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 7 contains the results, restricting our sample to consecutive observations during the period where the survey was fielded monthly. Table D.1 in the Online Appendix repeats the exercise for our entire sample with very similar results.

Table 7: Predictability of forecast errors with forecast revisions

	Pooled OLS	Pooled OLS w groups
Forecast Revision	-0.52	
	(0.00)	
Forecast Revision, Pessimists		-0.52
		(0.01)
Forecast Revision, Mean Reverters		-0.51
		(0.01)
Forecast Revision, Extrapolators		-0.52
		(0.01)
Forecast Revision, Ignorants		-0.51
		(0.01)
Forecast Revision, Sophisticates		-0.53
		(0.01)
R^2	0.12	0.30
N	50532	50532

OLS estimates. Standard errors (clustered at individual level) in parentheses.

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 future returns[3]. It is unsurprising that this is the case. If a weak form of the efficient market

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

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

hypothesis holds, then 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. If the weak form of the efficient market hypothesis holds, any response to such information is an overreaction. The results of Table 7 bear this out.

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 9% 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 both for literacy indices and for inconsistencies in the belief elicitation procedure) 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. 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. An important step will be, of course, to replicate our findings in other datasets. Whereas the groups we could identify were fairly stable in our setting, it would be important to see whether comparable findings emerge from other question formats and in other countries. If so, this would 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. Our results 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.

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.

Coibion, Olivier and Yuriy Gorodnichenko (2012). "What Can Survey Forecasts Tell Us about Information Rigidities?" In: *Journal of Political Economy* 120.1, pp. 116–159.

Dominitz, Jeff and Charles F. Manski (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.

Drerup, Tilman, Benjamin Enke, and Hans-Martin von Gaudecker (2017). "The Precision of Subjective Data and the Explanatory Power of Economic Models". In: *Journal of Econometrics*. Measurement Error Models 200.2, pp. 378–389.

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

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`.

Gouret, Fabian and Guillaume Hollard (2011). "When Kahneman Meets Manski: Using Dual Systems of Reasoning to Interpret Subjective Expectations of Equity Returns". In: *Journal of Applied Econometrics* 26.3, pp. 371–392.

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

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

Hudomiet, Péter, Gabor Kezdi, and Robert J. Willis (2010). "Stock market crash and expectations of American households". In: *Journal of Applied Econometrics*.

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.