Online appendix accompanying:

Heterogeneity in households' stock market beliefs *

Hans-Martin von Gaudecker and Axel Wogrolly

October 15, 2020

Contents

*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.

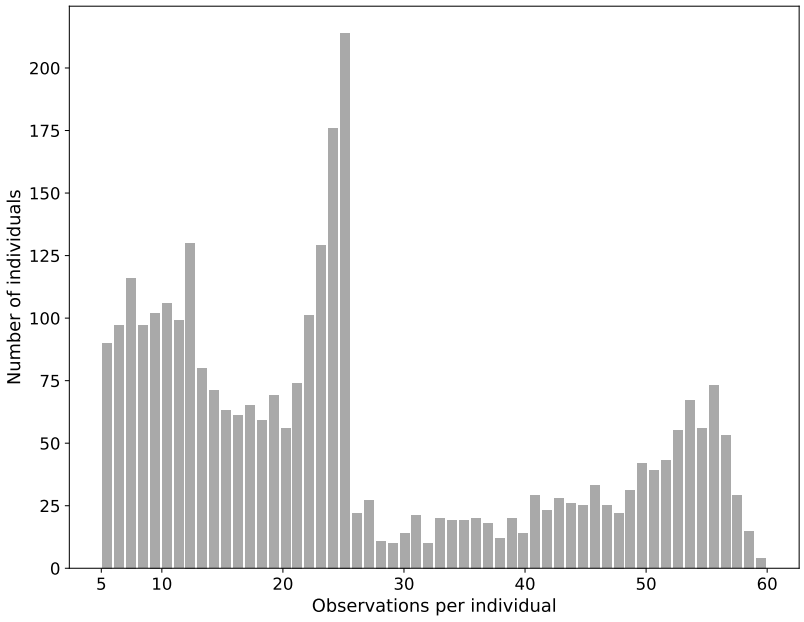
A Data and stylised facts

A.1 Additional details on the data and belief measures

We use several background variables from the ALP, which are available for everyone. These include age, sex, education, and income. The other variables were regularly measured as part of the "Effects of the Financial Crisis" survey waves, or they come from other surveys in the ALP. A detailed source for each variable is given below so all details can be retrieved from `https://alpdata.rand.org/index.php?page=data`. The main differences are that we combine the two survey identifiers to be found for waves 16 (ALP survey identifiers 129 and 131) and 44 (288, 293) of the Effects of the Financial Crisis survey and that we display the number of observations for each variable that we can effectively use. The number of belief measures per wave is reduced substantially midway through the sample because a second and hard-to-compare format for belief measurement was introduced; which format was shown to individuals was drawn randomly anew in each wave.

Figure A.3 shows histograms of the beliefs with 1-percent bins. Most beliefs are rounded to the nearest multiple of 5%, many to 10%, and answers equalling 50% are particularly frequent. The middle Panel of Figure A.3 looks very similar to Figure 3 in Hurd, Rooij, and

Figure A.1: Distribution of the number of belief measurements per individual



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.

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 A.7 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 A.1: 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

N = 3030. Table shows fraction of beliefs satisfying each listed monotonicity status.

Table A.7 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.

A.2 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 A.2: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 10%
Follows stock market	-0.05**	-0.02	0.02
	(0.02)	(0.01)	(0.02)
Understands stock market	-0.03	-0.02**	-0.02
	(0.02)	(0.01)	(0.02)
Knowledge of past returns: Don't know	-0.07**	0.01	-0.01
	(0.03)	(0.02)	(0.03)
Knowledge of past returns: Magnitude too large	-0.05	0.01	-0.02
	(0.04)	(0.03)	(0.04)
Knowledge of past returns: Sign and Magnitude correct	-0.08***	-0.01	-0.04
	(0.03)	(0.02)	(0.03)
Probability Numeracy	-0.06***	-0.01	0.01
	(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	0.00
	(0.01)	(0.01)	(0.01)
Intercept	0.59***	0.18	0.66***
	(0.17)	(0.11)	(0.22)
Age	0.01***	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Age squared	-0.00***	0.00	0.00
	(0.00)	(0.00)	(0.00)
Male	-0.02	0.00	-0.01
	(0.02)	(0.01)	(0.02)
Education: Some college	0.01	0.01	0.02
	(0.03)	(0.02)	(0.03)
Education: Bachelor degree	-0.05*	0.01	0.04
	(0.03)	(0.02)	(0.03)
Education: Advanced degree	-0.06*	-0.01	-0.02
	(0.03)	(0.02)	(0.03)
Ethnicity: Black	0.04	-0.02	0.10
	(0.08)	(0.04)	(0.07)
Ethnicity: Native	0.31***	-0.11***	-0.01
	(0.08)	(0.04)	(0.07)
Ethnicity: Other	0.00	0.00	0.17
	(0.11)	(0.07)	(0.11)
Ethnicity: White	-0.10	-0.01	0.08
	(0.07)	(0.03)	(0.05)
Household income (thousands), $\in (5, 7.5]$	-0.09	-0.01	0.18
	(0.15)	(0.08)	(0.20)
Household income (thousands), $\in (7.5, 10]$	-0.07	0.01	0.07
	(0.11)	(0.10)	(0.17)
Household income (thousands), $\in (10, 12.5]$	0.09	0.05	0.11
	(0.11)	(0.09)	(0.18)
Household income (thousands), $\in (12.5, 15]$	-0.05	-0.02	-0.02
	(0.11)	(0.08)	(0.17)
Household income (thousands), $\in (15, 20]$	0.00	-0.03	0.08
	(0.11)	(0.08)	(0.17)

Continued on next page

Table A.2: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 10%
Household income (thousands), $\in (20, 25]$	-0.01	0.04	0.13
	(0.10)	(0.08)	(0.16)
Household income (thousands), $\in (25, 30]$	-0.05	-0.05	-0.00
	(0.10)	(0.07)	(0.16)
Household income (thousands), $\in (30, 35]$	0.01	0.02	0.09
	(0.10)	(0.08)	(0.16)
Household income (thousands), $\in (35, 40]$	-0.09	-0.04	0.03
	(0.10)	(0.07)	(0.16)
Household income (thousands), $\in (40, 50]$	-0.03	-0.00	0.08
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (50, 60]$	-0.07	0.02	0.09
	(0.09)	(0.08)	(0.16)
Household income (thousands), $\in (60, 75]$	-0.02	-0.01	0.06
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (75, 100]$	-0.09	-0.01	0.06
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (100, 125]$	-0.08	-0.03	0.02
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (125, 200]$	-0.10	-0.03	0.10
	(0.09)	(0.07)	(0.16)
Household income (thousands), > 200	-0.08	-0.02	0.13
	(0.10)	(0.08)	(0.16)
Ever owned stocks	0.07***	0.04***	0.04*
	(0.02)	(0.01)	(0.02)
$\Pr(R_{t\to t+12} > 0)$	0.49***	0.11***	-0.34***
	(0.05)	(0.03)	(0.06)
N	805	805	805
R^2	0.31	0.13	0.13

Individuals for which not all covariates are available are excluded. 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.

In line with the earlier discussion, the regression results in Table A.8 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 10, is not systematically predictable with our indicators of sophistication.

A.3 Additional details

We use several background variables from the ALP, which are available for everyone. These include age, sex, education, and income. The other variables were regularly measured as part of the "Effects of the Financial Crisis" survey waves, or they come from other surveys in the ALP. A detailed source for each variable is given below so all details can be retrieved from `https://alpdata.rand.org/index.php?page=data`. The main differences are that we combine the two survey identifiers to be found for waves 16 (ALP survey identifiers 129 and 131) and 44 (288, 293) of the Effects of the Financial Crisis survey and that we display the number of observations for each variable that we can effectively use. The number of belief measures per wave is reduced substantially midway through the sample because a second and hard-to-compare format for belief measurement was introduced; which format was shown to individuals was drawn randomly anew in each wave.

Table A.3: Details on surveys and variables

ID	Survey Title	Fielded	Variable Name	N
57	Effects of the Financial Crisis [W01]	2008-11	Beliefs	1840
			Follows/Understands stock market	1867
63	Effects of the Financial Crisis [W02]	2009-02	Beliefs	1710
			Follows/Understands stock market	1970
64	Financial Literacy March 09	2009-03	Financial Numeracy	1564
			Financial Knowledge	
74	Effects of the Financial Crisis [W03]	2009-05	Beliefs	1796
			Owns stocks	1978
83	Effects of the Financial Crisis [W04]	2009-06	Beliefs	1876
			Owns stocks	2017
85	Effects of the Financial Crisis [W05]	2009-07	Beliefs	1904
			Owns stocks	2034
88	Effects of the Financial Crisis [W06]	2009-08	Beliefs	1867
			Owns stocks	2026
90	Effects of the Financial Crisis [W07]	2009-09	Beliefs	1901
			Owns stocks	2060
92	Effects of the Financial Crisis [W08]	2009-10	Beliefs	1816
			Owns stocks	1971
97	Effects of the Financial Crisis [W09]	2009-11	Beliefs	1891
			Owns stocks	2022
103	Effects of the Financial Crisis [W10]	2009-12	Beliefs	1904

Continued on next page

ID	Survey Title	Fielded	Variable Name	N
			Owns stocks	2036
107	Effects of the Financial Crisis [W11]	2010-01	Beliefs	1879
			Owns stocks	2032
			Follows/Understands stock market	2040
111	Effects of the Financial Crisis [W12]	2010-02	Beliefs	1890
			Owns stocks	2023
116	Effects of the Financial Crisis [W13]	2010-03	Beliefs	1862
			Owns stocks	1996
117	Effects of the Financial Crisis [W14]	2010-04	Beliefs	1798
			Owns stocks	1927
			Follows/Understands stock market	1934
124	Effects of the Financial Crisis [W15]	2010-05	Beliefs	1720
			Owns stocks	1852
129	Effects of the Financial Crisis [W16]	2010-06	Beliefs	1775
			Owns stocks	1911
134	Effects of the Financial Crisis [W17]	2010-07	Beliefs	1668
			Owns stocks	1793
139	Effects of the Financial Crisis [W18]	2010-08	Beliefs	1640
			Owns stocks	1741
152	Effects of the Financial Crisis [W19]	2010-09	Beliefs	1695
			Owns stocks	1817
157	Effects of the Financial Crisis [W20]	2010-10	Beliefs	1659
			Owns stocks	1770
158	Effects of the Financial Crisis [W21]	2010-11	Beliefs	1706
			Owns stocks	1830
161	Effects of the Financial Crisis [W22]	2010-12	Beliefs	1726
			Owns stocks	1848
162	Effects of the Financial Crisis [W23]	2011-01	Beliefs	1666
			Owns stocks	1801
173	Effects of the Financial Crisis [W24]	2011-02	Beliefs	1722
			Owns stocks	1812
			Follows/Understands stock market	1815
176	Effects of the Financial Crisis [W25]	2011-03	Beliefs	1708
			Owns stocks	1828
178	Effects of the Financial Crisis [W26]	2011-04	Beliefs	860
			Owns stocks	1759
188	Effects of the Financial Crisis [W27]	2011-05	Beliefs	859
			Owns stocks	1748
194	Effects of the Financial Crisis [W28]	2011-06	Beliefs	872
			Owns stocks	1749
198	Effects of the Financial Crisis [W29]	2011-07	Beliefs	889
			Owns stocks	1805
208	Effects of the Financial Crisis [W30]	2011-08	Beliefs	883
			Owns stocks	1820
211	Effects of the Financial Crisis [W31]	2011-09	Beliefs	846
			Owns stocks	1759
219	Effects of the Financial Crisis [W32]	2011-10	Beliefs	826
			Owns stocks	1746
225	Effects of the Financial Crisis [W33]	2011-11	Beliefs	937
			Owns stocks	1792
231	Effects of the Financial Crisis [W34]	2011-12	Beliefs	898

Continued on next page

ID	Survey Title	Fielded	Variable Name	N
			Owns stocks	1742
236	Effects of the Financial Crisis [W35]	2012-01	Beliefs	960
			Owns stocks	1796
239	Effects of the Financial Crisis [W36]	2012-02	Beliefs	955
			Owns stocks	1795
249	Effects of the Financial Crisis [W37]	2012-03	Beliefs	906
			Owns stocks	1682
253	Effects of the Financial Crisis [W38]	2012-04	Beliefs	946
			Owns stocks	1756
			Follows/Understands stock market	1757
262	Effects of the Financial Crisis [W39]	2012-05	Beliefs	805
			Owns stocks	1507
267	Effects of the Financial Crisis [W40]	2012-06	Beliefs	892
			Owns stocks	1677
271	Effects of the Financial Crisis [W41]	2012-07	Beliefs	928
			Owns stocks	1743
			Follows/Understands stock market	1754
278	Effects of the Financial Crisis [W42]	2012-08	Beliefs	927
			Owns stocks	1698
281	Effects of the Financial Crisis [W43]	2012-09	Beliefs	907
			Owns stocks	1669
288	Effects of the Financial Crisis [W44]	2012-10	Beliefs	1023
			Owns stocks	1820
299	Effects of the Financial Crisis [W45]	2012-11	Beliefs	1312
			Owns stocks	2181
305	Effects of the Financial Crisis [W46]	2012-12	Beliefs	1347
			Owns stocks	2204
			Knows stock return from last year	1039
322	Effects of the Financial Crisis [W47]	2013-01	Beliefs	1101
			Owns stocks	1720
328	Effects of the Financial Crisis [W48]	2013-02	Beliefs	1317
			Owns stocks	2086
			Knows stock return from last year	1008
332	Effects of the Financial Crisis [W49]	2013-03	Beliefs	1300
			Owns stocks	2145
335	Effects of the Financial Crisis [W50]	2013-04	Beliefs	1347
			Owns stocks	2100
			Follows/Understands stock market	1467
			Knows stock return from last year	1176
345	Effects of the Financial Crisis [W51]	2013-07	Beliefs	98
			Owns stocks	157
358	Effects of the Financial Crisis [W52]	2013-10	Beliefs	881
			Owns stocks	1428
363	Reasons for expectations [W01]	2013-12	Reasons for expectations	114
368	Effects of the Financial Crisis [W53]	2014-01	Beliefs	939
			Owns stocks	1478
379	Effects of the Financial Crisis [W54]	2014-04	Beliefs	257
			Owns stocks	439

Continued on next page

ID	Survey Title	Fielded	Variable Name	N
			Follows/Understands stock market	279
389	Effects of the Financial Crisis [W55]	2014-07	Beliefs	924
			Owns stocks	1500
			Follows/Understands stock market	1498
400	Effects of the Financial Crisis [W56]	2014-10	Beliefs	311
			Owns stocks	540
417	Effects of the Financial Crisis [W57]	2015-01	Beliefs	949
			Owns stocks	1547
426	Effects of the Financial Crisis [W58]	2015-04	Beliefs	794
			Owns stocks	1303
			Follows/Understands stock market	1296
			Probability Numeracy	1291
434	Effects of the Financial Crisis [W59]	2015-07	Beliefs	825
			Owns stocks	1323
			Follows/Understands stock market	1320
440	Effects of the Financial Crisis [W60]	2015-10	Beliefs	813
			Owns stocks	1340
			Follows/Understands stock market	1335
			Probability Numeracy	1310
448	Effects of the Financial Crisis [W61]	2016-01	Beliefs	1057
			Owns stocks	1700
			Follows/Understands stock market	1693
			Probability Numeracy	1670

A.4 Detailed stylised facts (Section 2.2)

A.4.1 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 report for Dutch households.

In Table A.4 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 A.4: 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.

A.4.2 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.5).

Table A.5: Within- and between-person variation of belief variables

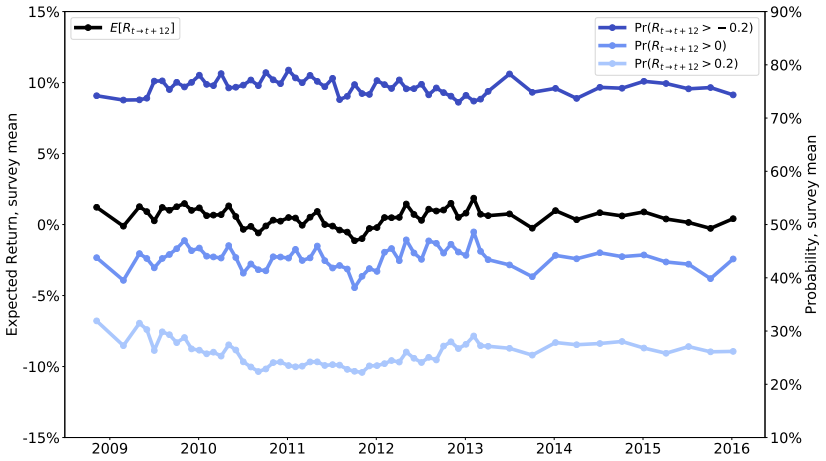
	Average within-subject std.	Between-subject std.	Ratio
$E[R_{t\to t+12}]$	5.5	5.8	0.94
$\Pr(R_{t\to t+12}) > -0.2$	13.3	13.4	1.00
$\Pr(R_{t\to t+12}) > 0$	15.5	17.8	0.87
$\Pr(R_{t\to t+12}) > 0.2$	14.6	14.2	1.03

Units in percentage points.

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 the next Subsection A.4.3, 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 A.2 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 A.2: 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.

A.4.3 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 A.6 reports the results.

Table A.6: Predictors of average beliefs, all regressors

	$E[R]$	$\Pr(R > -0.2)$	$\Pr(R > 0)$	$\Pr(R > 0.2)$
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)
Intercept	-10.25***	95.85***	-6.74	11.85
	(3.80)	(9.76)	(12.72)	(10.87)
Age	-0.02	-0.94***	0.20	0.36
	(0.09)	(0.23)	(0.28)	(0.25)
Age squared	0.00	0.01***	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Male	0.01	-1.26	2.21*	-1.45
	(0.43)	(0.95)	(1.25)	(1.08)
Education: Some college	0.69	-2.72*	2.15	2.88*
	(0.60)	(1.56)	(1.87)	(1.56)
Education: Bachelor degree	1.69***	-0.96	5.75***	3.10*
	(0.65)	(1.67)	(1.97)	(1.66)
Education: Advanced degree	2.71***	0.71	8.68***	3.47*
	(0.73)	(1.63)	(2.16)	(1.81)
Ethnicity: Black	1.32	1.43	4.21	1.50
	(1.80)	(3.62)	(4.92)	(5.32)
Ethnicity: Native	-0.36	-28.22***	10.16**	4.08
	(1.76)	(3.68)	(4.86)	(5.14)
Ethnicity: Other	-1.30	-4.70	0.02	-2.94
	(1.87)	(6.59)	(5.71)	(5.43)
Ethnicity: White	0.47	0.02	4.68	-2.28
	(1.52)	(3.03)	(3.98)	(4.68)
Household income (thousands), $\in (5, 7.5]$	3.04	-6.33	9.79	9.55
	(4.55)	(9.24)	(16.47)	(10.74)
Household income (thousands), $\in (7.5, 10]$	-0.28	0.50	0.99	-2.69

Continued on next page

Table A.6: Predictors of average beliefs, all regressors

	$E[R]$	$\Pr(R > -0.2)$	$\Pr(R > 0)$	$\Pr(R > 0.2)$
	(3.08)	(7.96)	(10.74)	(7.98)
Household income (thousands), $\in (10, 12.5]$	-1.86	-12.31	-1.27	2.30
	(3.36)	(8.31)	(10.93)	(8.04)
Household income (thousands), $\in (12.5, 15]$	0.25	-0.88	1.65	0.39
	(3.03)	(7.38)	(11.15)	(7.82)
Household income (thousands), $\in (15, 20]$	1.24	-5.98	4.04	6.73
	(3.00)	(7.77)	(10.86)	(7.86)
Household income (thousands), $\in (20, 25]$	0.85	-6.14	3.15	5.15
	(2.89)	(7.15)	(10.54)	(7.67)
Household income (thousands), $\in (25, 30]$	0.23	-0.23	1.28	0.61
	(2.89)	(7.26)	(10.76)	(7.58)
Household income (thousands), $\in (30, 35]$	0.32	-4.00	1.07	3.99
	(2.86)	(7.04)	(10.51)	(7.47)
Household income (thousands), $\in (35, 40]$	2.50	-0.97	7.33	5.29
	(2.84)	(7.03)	(10.50)	(7.48)
Household income (thousands), $\in (40, 50]$	0.64	-3.00	3.41	2.12
	(2.79)	(6.88)	(10.31)	(7.32)
Household income (thousands), $\in (50, 60]$	1.13	-2.39	4.74	2.90
	(2.82)	(6.93)	(10.37)	(7.42)
Household income (thousands), $\in (60, 75]$	0.93	-3.16	3.00	4.15
	(2.81)	(6.85)	(10.33)	(7.34)
Household income (thousands), $\in (75, 100]$	1.73	-0.83	5.49	3.40
	(2.78)	(6.80)	(10.29)	(7.25)
Household income (thousands), $\in (100, 125]$	0.64	-2.41	3.98	0.97
	(2.79)	(6.81)	(10.34)	(7.28)
Household income (thousands), $\in (125, 200]$	0.76	-0.57	2.21	2.24
	(2.84)	(6.82)	(10.39)	(7.40)
Household income (thousands), $> 200k$	-0.92	-2.17	-1.78	-0.45
	(3.01)	(7.15)	(10.75)	(7.68)
Ever owned stocks	-0.29	-1.16	-1.03	0.86
	(0.57)	(1.21)	(1.57)	(1.37)
N	805	805	805	805
R^2	0.22	0.11	0.32	0.05

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.

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.

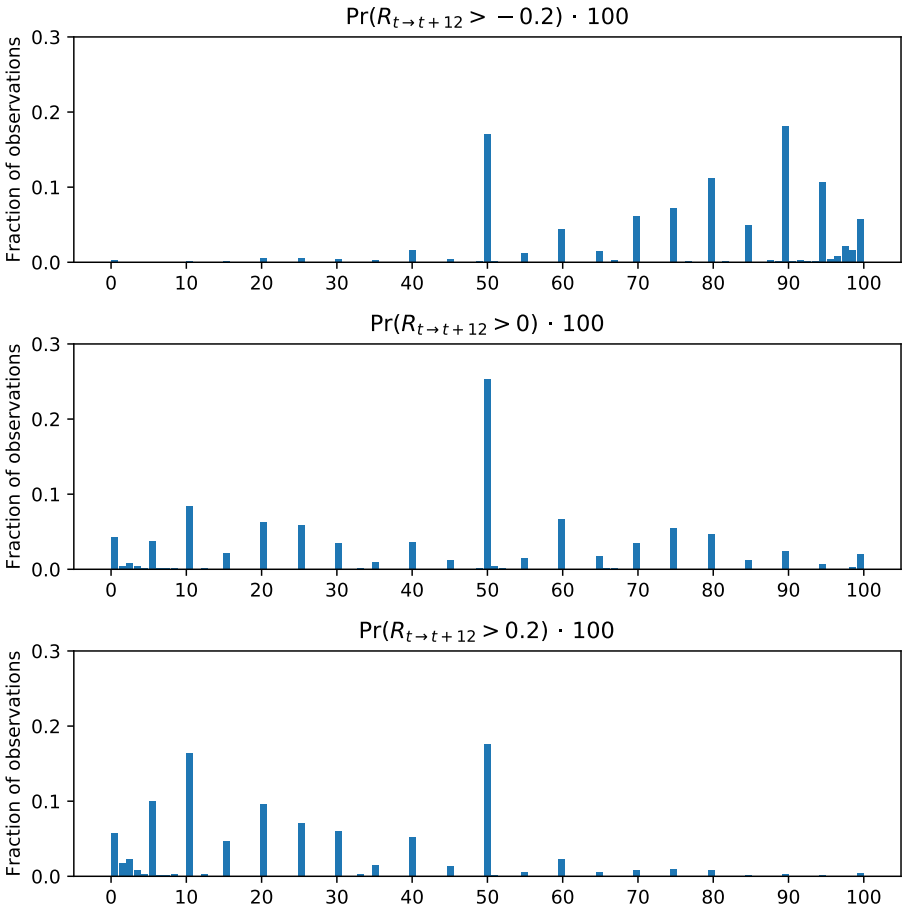
As noted before, the predictive power of the covariates is much higher for the probability of a positive return than for the other two points on the distribution function; the R^2 differs by factors of three to five. We take this as additional evidence pointing towards higher noise levels for the events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

A.4.4 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 A.3 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 A.3 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.

Figure A.3: 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.

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 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.

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 A.7 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 A.7: 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 A.7 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.

A.4.5 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 A.8: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 10%
Follows stock market	-0.05**	-0.02	0.02
	(0.02)	(0.01)	(0.02)
Understands stock market	-0.03	-0.02**	-0.02
	(0.02)	(0.01)	(0.02)
Knowledge of past returns: Don't know	-0.07**	0.01	-0.01
	(0.03)	(0.02)	(0.03)
Knowledge of past returns: Magnitude too large	-0.05	0.01	-0.02
	(0.04)	(0.03)	(0.04)
Knowledge of past returns: Sign and Magnitude correct	-0.08***	-0.01	-0.04
	(0.03)	(0.02)	(0.03)
Probability Numeracy	-0.06***	-0.01	0.01
	(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	0.00
	(0.01)	(0.01)	(0.01)
Intercept	0.59***	0.18	0.66***
	(0.17)	(0.11)	(0.22)

Continued on next page

Table A.8: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 10%
Age	0.01***	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Age squared	-0.00***	0.00	0.00
	(0.00)	(0.00)	(0.00)
Male	-0.02	0.00	-0.01
	(0.02)	(0.01)	(0.02)
Education: Some college	0.01	0.01	0.02
	(0.03)	(0.02)	(0.03)
Education: Bachelor degree	-0.05*	0.01	0.04
	(0.03)	(0.02)	(0.03)
Education: Advanced degree	-0.06*	-0.01	-0.02
	(0.03)	(0.02)	(0.03)
Ethnicity: Black	0.04	-0.02	0.10
	(0.08)	(0.04)	(0.07)
Ethnicity: Native	0.31***	-0.11***	-0.01
	(0.08)	(0.04)	(0.07)
Ethnicity: Other	0.00	0.00	0.17
	(0.11)	(0.07)	(0.11)
Ethnicity: White	-0.10	-0.01	0.08
	(0.07)	(0.03)	(0.05)
Household income (thousands), $\in (5, 7.5]$	-0.09	-0.01	0.18
	(0.15)	(0.08)	(0.20)
Household income (thousands), $\in (7.5, 10]$	-0.07	0.01	0.07
	(0.11)	(0.10)	(0.17)
Household income (thousands), $\in (10, 12.5]$	0.09	0.05	0.11
	(0.11)	(0.09)	(0.18)
Household income (thousands), $\in (12.5, 15]$	-0.05	-0.02	-0.02
	(0.11)	(0.08)	(0.17)
Household income (thousands), $\in (15, 20]$	0.00	-0.03	0.08
	(0.11)	(0.08)	(0.17)
Household income (thousands), $\in (20, 25]$	-0.01	0.04	0.13
	(0.10)	(0.08)	(0.16)
Household income (thousands), $\in (25, 30]$	-0.05	-0.05	-0.00
	(0.10)	(0.07)	(0.16)
Household income (thousands), $\in (30, 35]$	0.01	0.02	0.09
	(0.10)	(0.08)	(0.16)
Household income (thousands), $\in (35, 40]$	-0.09	-0.04	0.03
	(0.10)	(0.07)	(0.16)
Household income (thousands), $\in (40, 50]$	-0.03	-0.00	0.08
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (50, 60]$	-0.07	0.02	0.09
	(0.09)	(0.08)	(0.16)
Household income (thousands), $\in (60, 75]$	-0.02	-0.01	0.06
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (75, 100]$	-0.09	-0.01	0.06
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (100, 125]$	-0.08	-0.03	0.02
	(0.09)	(0.07)	(0.16)
Household income (thousands), $\in (125, 200]$	-0.10	-0.03	0.10
	(0.09)	(0.07)	(0.16)
Household income (thousands), > 200	-0.08	-0.02	0.13
	(0.10)	(0.08)	(0.16)

Continued on next page

Table A.8: Predictors of non-monotonicity, epistimic uncertainty, and rounding

	Non-monotone	Epistemically unsure	Rounded to 10%
Ever owned stocks	0.07***	0.04***	0.04*
	(0.02)	(0.01)	(0.02)
$\Pr(R_{t\to t+12} > 0)$	0.49***	0.11***	-0.34***
	(0.05)	(0.03)	(0.06)
N	805	805	805
R^2	0.31	0.13	0.13

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.

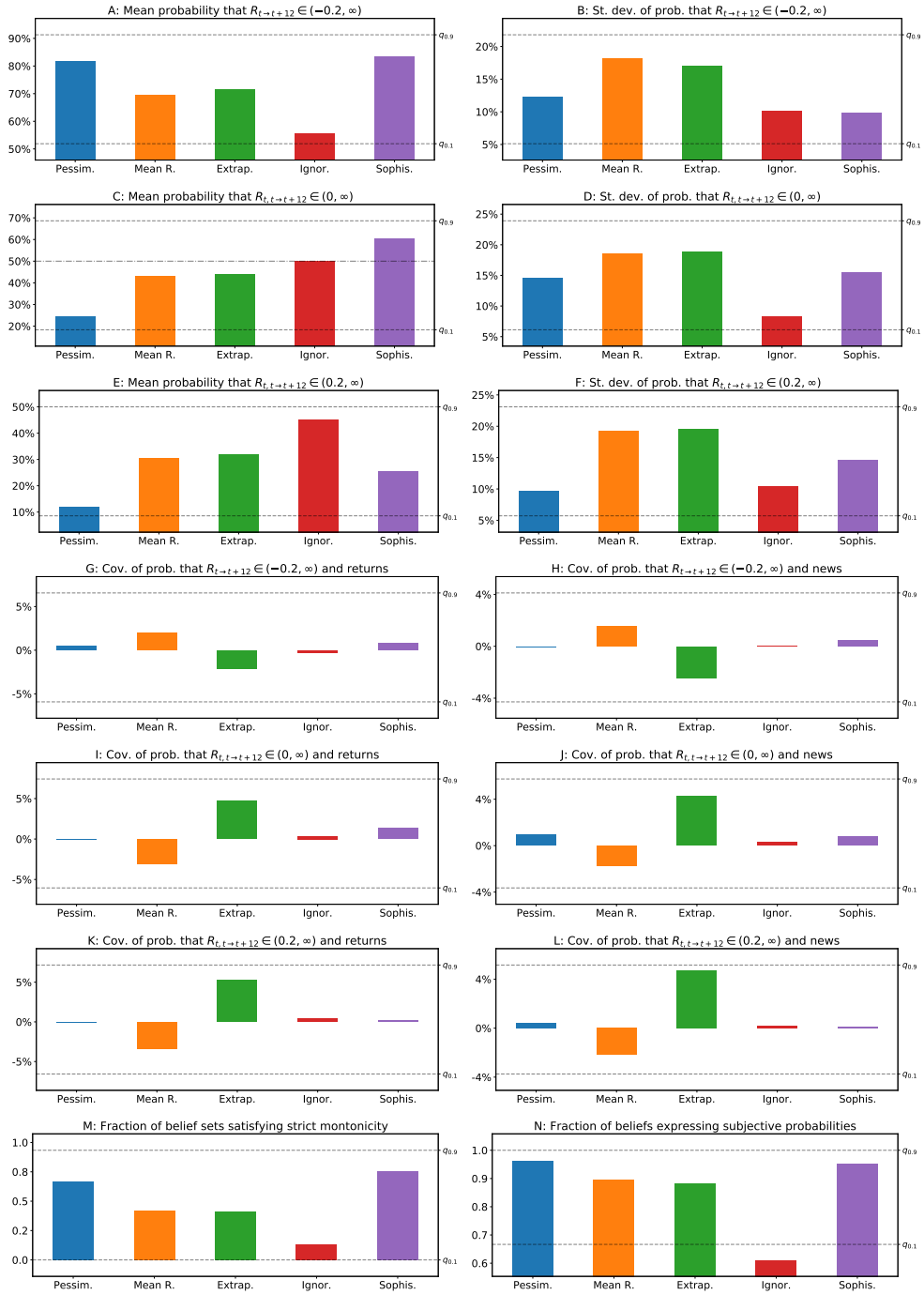
In line with the earlier discussion, the regression results in Table A.8 demonstrate that following the stock market, having accurate knowledge of historical returns, probability numeracy, and financial numeracy all predict that an individual is less likely to express beliefs afflicted by monotonicity 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.

B Additional details for the main model

Table B.1: Group sizes

	N	Share of sample
Pessimists	749	0.25
Mean Reverters	584	0.19
Extrapolators	527	0.17
Ignorants	393	0.13
Sophisticates	777	0.26
Total	3030	

Figure B.1: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Table B.2: Coefficients for main specification when L = 0

	Pessim.	Mean R.	Extrap.	Ignor.	Sophis.
Intercept	-4.74***	-0.45	3.06***	2.93***	5.32***
	(0.15)	(0.32)	(0.32)	(0.21)	(0.18)
Lag 0, Returns	-0.02	-0.53***	0.72***	0.12**	0.24***
	(0.03)	(0.06)	(0.07)	(0.05)	(0.04)
Lag 0, News	0.42***	-0.47***	1.15***	0.07	0.27***
	(0.05)	(0.10)	(0.10)	(0.06)	(0.05)
N · T			77310		
R^2			0.256		

N = 3030. Individuals for whom not all covariates are available are excluded. OLS estimates. Standard errors (clustered at individual level) in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. Dependent variable in percentage points, regressors standardised.

Table B.3: Coefficients for main specification when L = 6

	Pessim.	Mean R.	Extrap.	Ignor.	Sophis.
Intercept	-3.35***	-0.42	3.73***	3.26***	6.22***
	(0.28)	(0.61)	(0.60)	(0.35)	(0.33)
Lag 0, Returns	0.04	-0.55***	0.62***	0.09*	0.22***
	(0.04)	(0.07)	(0.07)	(0.05)	(0.04)
Lag 1, Returns	-0.02	0.04	0.41***	0.17***	0.15***
	(0.04)	(0.07)	(0.07)	(0.05)	(0.04)
Lag 2, Returns	0.01	-0.09	-0.00	0.16***	0.10***
	(0.03)	(0.06)	(0.07)	(0.05)	(0.04)
Lag 3, Returns	-0.01	-0.05	0.11*	0.05	0.12***
	(0.03)	(0.06)	(0.07)	(0.04)	(0.03)
Lag 4, Returns	-0.08***	0.01	0.10*	0.06*	0.08**
	(0.03)	(0.05)	(0.06)	(0.04)	(0.03)
Lag 5, Returns	-0.08***	-0.10*	-0.01	0.08**	0.01
	(0.03)	(0.06)	(0.06)	(0.03)	(0.03)
Lag 6, Returns	-0.08***	0.05	0.00	-0.01	0.02
	(0.03)	(0.06)	(0.07)	(0.04)	(0.03)
Lag 0, News	0.32***	-0.45***	1.04***	0.12*	0.17***
	(0.05)	(0.10)	(0.09)	(0.06)	(0.05)
Lag 1, News	0.13***	-0.03	0.13	-0.26***	-0.02
	(0.05)	(0.11)	(0.10)	(0.07)	(0.05)
Lag 2, News	0.23***	0.33***	0.14	0.07	0.31***
	(0.05)	(0.12)	(0.12)	(0.06)	(0.06)
Lag 3, News	0.16***	-0.04	0.41***	0.07	0.06
	(0.05)	(0.12)	(0.13)	(0.08)	(0.06)
Lag 4, News	-0.03	-0.26**	-0.10	0.08	0.05
	(0.06)	(0.12)	(0.13)	(0.08)	(0.07)
Lag 5, News	0.07	0.10	-0.24*	0.06	0.04
	(0.06)	(0.13)	(0.13)	(0.08)	(0.07)
Lag 6, News	0.19***	-0.09	0.14	0.14*	0.13**
	(0.06)	(0.13)	(0.13)	(0.08)	(0.06)
$N \cdot T$			77310		
R^2			0.256		

N = 3030. OLS estimates. Standard errors (clustered at individual level) in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively. Dependent variable in percentage points, regressors standardised.

Table B.4: Within-group heterogeneity

		Main specifciation	Fin. Know.: > med	Fin. Num.: > med	Underst. stock m.	Follows stock m.	Know. past ret.	Age: > med
Pessim.	Returns	-0.04	0.31***	0.28***	0.30***	0.24**	0.25***	0.02
		(0.05)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.11)
	News	0.47***	-0.02	-0.25*	-0.16	-0.16	-0.16	0.10
		(0.07)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)	(0.16)
Mean R.	Returns	-0.52***	0.06	-0.05	0.05	-0.18	0.16	-0.05
		(0.09)	(0.20)	(0.19)	(0.19)	(0.18)	(0.19)	(0.20)
	News	-0.14	0.00	-0.21	-0.11	0.20	-0.31	0.26
		(0.15)	(0.33)	(0.30)	(0.32)	(0.27)	(0.30)	(0.37)
Extrap.	Returns	0.55***	0.45**	0.30	0.37*	0.63***	0.11	-0.01
		(0.11)	(0.20)	(0.22)	(0.22)	(0.20)	(0.21)	(0.24)
	News	1.14***	-0.30	0.20	-0.27	-0.18	-0.11	0.36
		(0.15)	(0.31)	(0.31)	(0.31)	(0.30)	(0.30)	(0.41)
Ignor.	Returns	0.11	-0.28	-0.01	-0.19	-0.34**	-0.04	-0.20
		(0.08)	(0.18)	(0.23)	(0.17)	(0.16)	(0.17)	(0.14)
	News	0.03	-0.03	-0.21	-0.08	0.01	-0.11	0.10
		(0.09)	(0.21)	(0.25)	(0.20)	(0.18)	(0.18)	(0.26)
Sophis.	Returns	0.22***	-0.03	-0.16	-0.02	0.05	0.08	-0.07
		(0.05)	(0.11)	(0.10)	(0.11)	(0.12)	(0.11)	(0.11)
	News	0.33***	-0.37***	-0.12	-0.29**	-0.28*	0.01	0.21
		(0.06)	(0.13)	(0.12)	(0.13)	(0.16)	(0.13)	(0.15)
	N · T	37828	37828	37828	37828	37828	37828	37828
	R^2	0.26	0.26	0.27	0.26	0.27	0.26	0.26

Individuals for whom not all covariates are available are excluded. The first column reproduces the coefficients from our main specification for the subsample of individuals for which all covariates in the adjacent columns are available. The adjacent columns show the difference for each row's coefficient between individuals with and without the status given in the column header.

Table B.5: Robustness to rounding

		Main specifciation	With indicators for rounding
Pessim.	Intercept	-4.74***	-3.94***
		(0.15)	(0.21)
	Returns	-0.02	-0.02
		(0.03)	(0.04)
	News	0.42***	0.42***
		(0.05)	(0.05)
Mean R.	Intercept	-0.45	0.27
		(0.32)	(0.37)
	Returns	-0.53***	-0.52***
		(0.06)	(0.06)
	News	-0.47***	-0.46***
		(0.10)	(0.09)
Extrap.	Intercept	3.06***	3.72***
		(0.32)	(0.37)
	Returns	0.72***	0.67***
		(0.07)	(0.07)
	News	1.15***	1.14***
		(0.10)	(0.10)
Ignor.	Intercept	2.93***	3.90***
		(0.21)	(0.31)
	Returns	0.12**	0.10**
		(0.05)	(0.05)
	News	0.07	0.09
		(0.06)	(0.06)
Sophis.	Intercept	5.32***	5.63***
		(0.18)	(0.21)
	Returns	0.24***	0.20***
		(0.04)	(0.04)
	News	0.27***	0.24***
		(0.05)	(0.05)
	$\Pr(R \leq -20\%)$ divisible by 10%		0.54***
			(0.11)
	$\Pr(R \leq 0\%)$ divisible by 10%		-3.33***
			(0.16)
	$\Pr(R \leq 20\%)$ divisible by 10%		1.62***
			(0.11)
	N · T	77310	77310
	R^2	0.25	0.29

The divisibility variables are dummies equal to 1 if the subjective probability it refers to is divisible by 10%. The outcome of the regressions, the expected return, is based on all three subjective probabilities.

C Alternative specifications

C.1 Only observed heterogeneity

Table C.1: Model with no heterogeneity and observed heterogeneity

	$E[R]$, no heterogeneity	$E[R]$, obs. heterogeneity
Intercept	1.83***	-6.80*
	(0.21)	(3.70)
Lag 0, Returns	0.06**	0.03
	(0.03)	(0.04)
Lag 1, Returns	0.12***	0.08**
	(0.03)	(0.04)
Lag 2, Returns	0.02	0.05*
	(0.02)	(0.03)
Lag 3, Returns	0.04*	0.08**
	(0.02)	(0.03)
Lag 4, Returns	0.02	0.04
	(0.02)	(0.03)
Lag 5, Returns	-0.02	-0.01
	(0.02)	(0.03)
Lag 6, Returns	0.00	0.02
	(0.02)	(0.03)
Lag 0, News	0.25***	0.31***
	(0.03)	(0.05)
Lag 1, News	-0.03	-0.03
	(0.03)	(0.05)
Lag 2, News	0.22***	0.33***
	(0.04)	(0.05)
Lag 3, News	0.13***	0.08
	(0.04)	(0.06)
Lag 4, News	-0.05	-0.01
	(0.04)	(0.06)
Lag 5, News	0.00	-0.00
	(0.04)	(0.06)
Lag 6, News	0.12***	0.14**
	(0.04)	(0.06)
Age		-0.01
		(0.09)
Age squared		0.00
		(0.00)
Male		0.13
		(0.46)
Education: Some college		0.70
		(0.68)
Education: Bachelor degree		1.91***
		(0.71)
Education: Advanced degree		2.71***
		(0.77)
Ethnicity: Black		1.82
		(2.06)
Ethnicity: Native		0.17
		(1.97)

Continued on next page

Table C.1: Model with no heterogeneity and observed heterogeneity

	$E[R]$, no heterogeneity	$E[R]$, obs. heterogeneity
Ethnicity: Other		-1.08
		(2.13)
Ethnicity: White		0.22
		(1.79)
Household income (thousands), $\in (5, 7.5]$		0.50
		(4.26)
Household income (thousands), $\in (7.5, 10]$		-0.54
		(2.45)
Household income (thousands), $\in (10, 12.5]$		-3.08
		(3.15)
Household income (thousands), $\in (12.5, 15]$		-0.41
		(2.44)
Household income (thousands), $\in (15, 20]$		0.27
		(2.46)
Household income (thousands), $\in (20, 25]$		-0.21
		(2.44)
Household income (thousands), $\in (25, 30]$		-1.26
		(2.37)
Household income (thousands), $\in (30, 35]$		-0.62
		(2.29)
Household income (thousands), $\in (35, 40]$		1.69
		(2.30)
Household income (thousands), $\in (40, 50]$		-0.50
		(2.21)
Household income (thousands), $\in (50, 60]$		-0.38
		(2.24)
Household income (thousands), $\in (60, 75]$		-0.30
		(2.23)
Household income (thousands), > 75 (higher cat n/a)		-1.67
		(2.24)
Household income (thousands), $\in (75, 100]$		0.22
		(2.18)
Household income (thousands), $\in (100, 125]$		-0.77
		(2.21)
Household income (thousands), $\in (125, 200]$		-0.21
		(2.25)
Household income (thousands), $> 200k$		-1.93
		(2.39)
Ever owned stocks		-0.09
		(0.61)
Follows stock market		1.18**
		(0.56)
Understands stock market		0.07
		(0.52)
Knowledge of past returns: Don't know		1.24*
		(0.64)
Knowledge of past returns: Magnitude too large		5.07***
		(1.06)
Knowledge of past returns: Sign and Magnitude correct		3.22***
		(0.66)
Probability Numeracy		0.98***
		(0.28)

Continued on next page

Table C.1: Model with no heterogeneity and observed heterogeneity

	$E[R]$, no heterogeneity	$E[R]$, obs. heterogeneity
Financial Knowledge		0.29
		(0.26)
Financial Numeracy		0.19
		(0.26)
N · T	77310	32170
R^2	0.00	0.12

N = 3030 individuals for the model of the first column. Requiring all covariates drops the number of individuals to N = 806 in the second column. OLS estimates. Standard errors (clustered by individual) in parentheses. ***, ** and * denote significance at 1%, 5% and 10% respectively.

C.2 Sample with at least three observations per individuals

This section shows the results when we include individuals with at least 2 observations into the analysis, which is the minimum needed to calculate all the variables by which we classify individuals into groups.

Table C.2: Group sizes

group	N	Share of sample
0	822	0.25
1	417	0.13
2	778	0.23
3	424	0.13
4	887	0.27
Total	3328	

Figure C.1: unit-wise 90% confidence sets by size and inclusion of estimated group

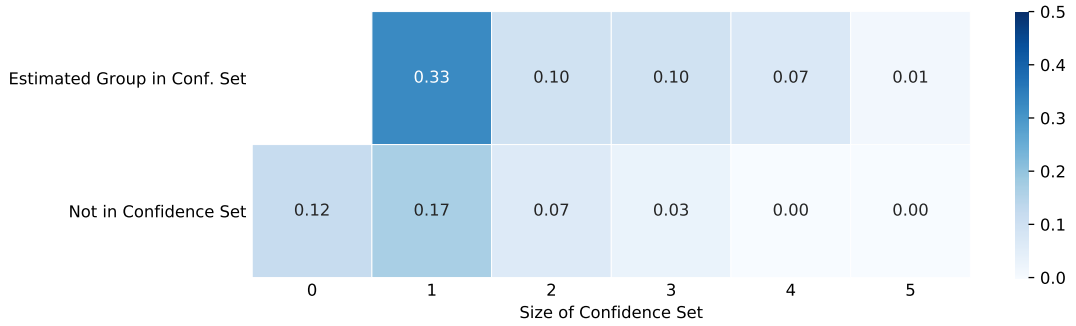
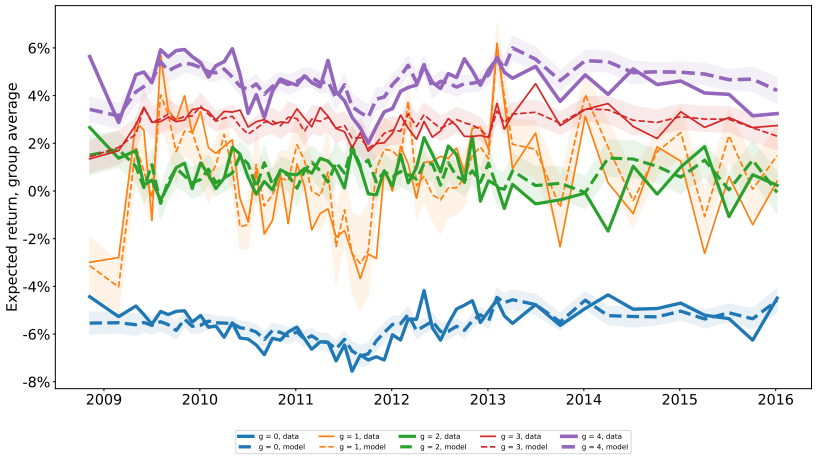


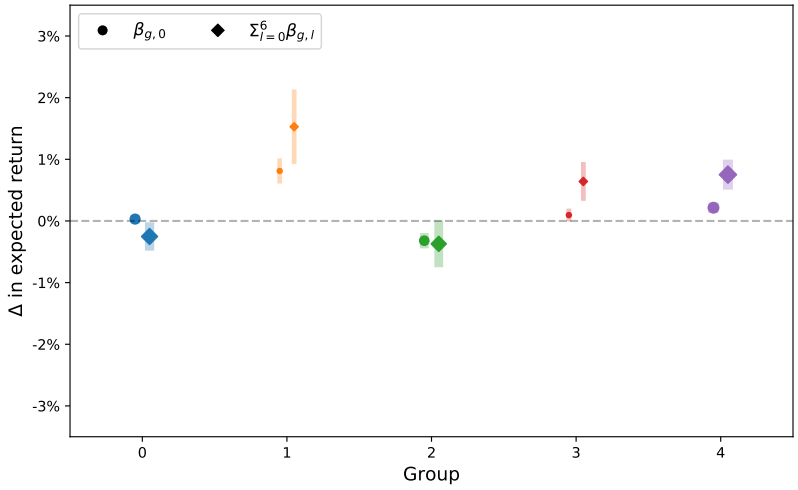
Figure C.2: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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. Where within survey and group means consist of less than 15 observations, we do not plot the series, resulting in a gap. Some ALP surveys had a smaller number of individuals taking part.

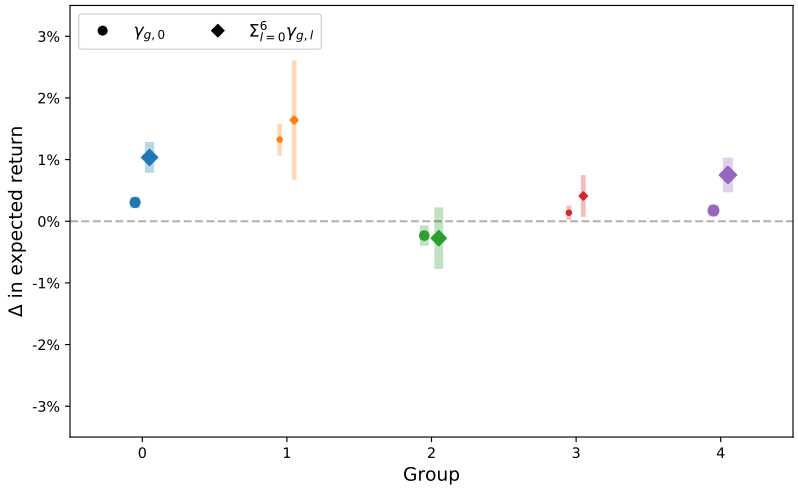
Figure C.3: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



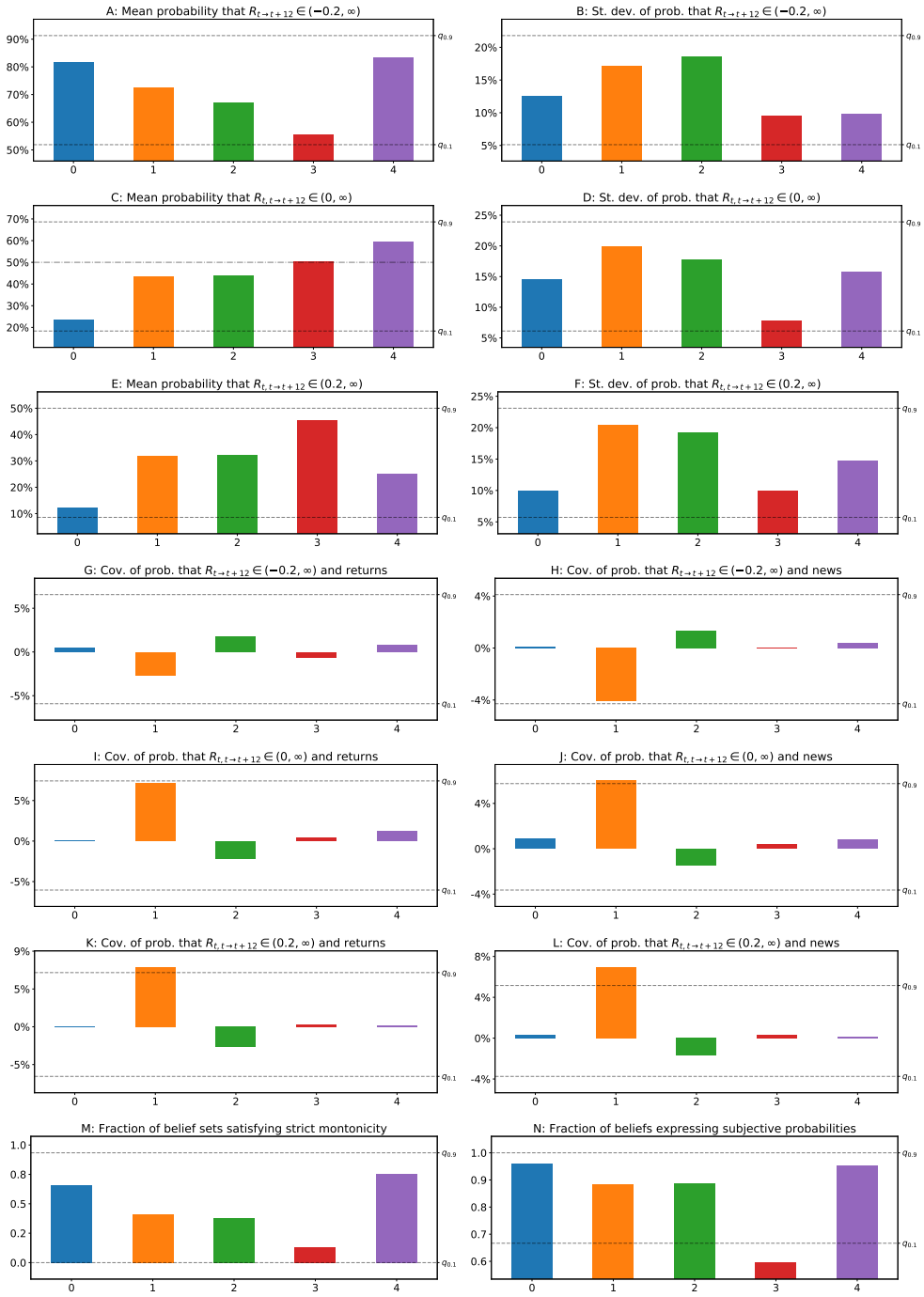
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



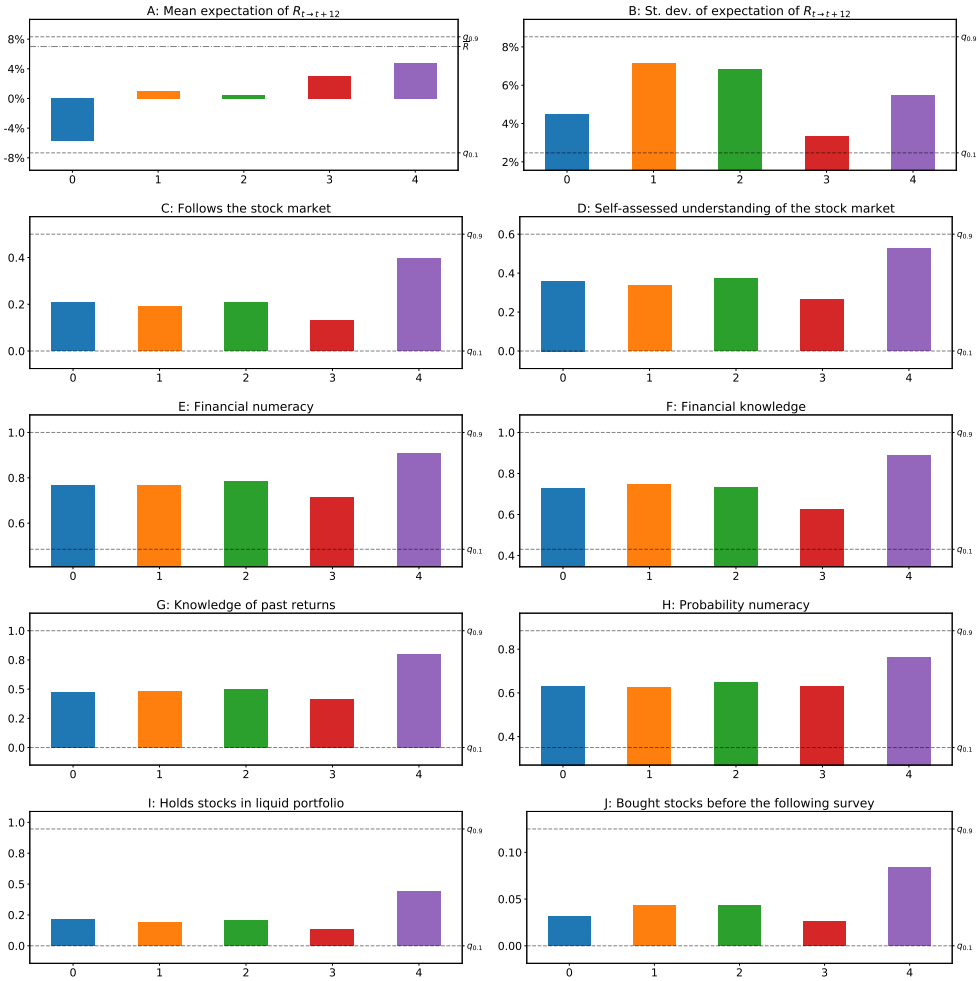
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.4: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.5: Characteristics not used for classification by unobserved heterogeneity group



N = 3328, 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).

Table C.3: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2	3	4
Pessimists	0.97	0.00	0.01	0.00	0.02
Mean Reverters	0.05	0.00	0.90	0.00	0.05
Extrapolators	0.04	0.65	0.21	0.03	0.08
Ignorants	0.00	0.00	0.08	0.91	0.00
Sophisticates	0.00	0.00	0.02	0.00	0.98

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

C.3 Sample with at least fifteen observations per individuals

This section shows the results when we include only individuals with at least 15 observations into the analysis, which is the number of parameters in our model when we include 6 lags of returns and news.

Table C.4: Group sizes

group	N	Share of sample
0	475	0.23
1	367	0.18
2	410	0.20
3	264	0.13
4	526	0.26
Total	2042	

Figure C.6: unit-wise 90% confidence sets by size and inclusion of estimated group

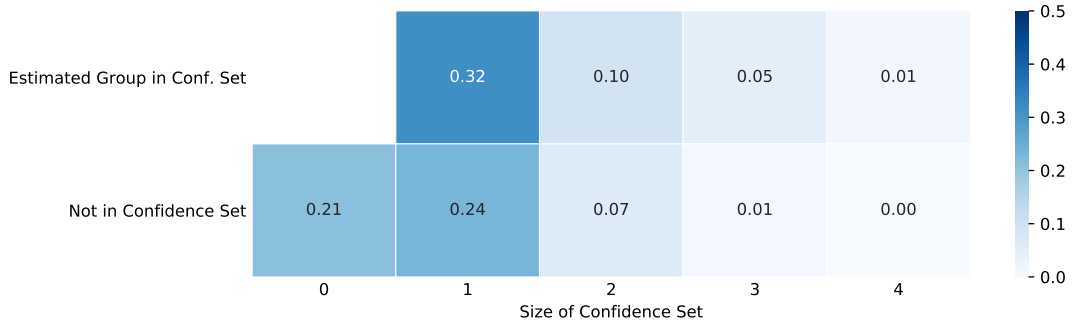
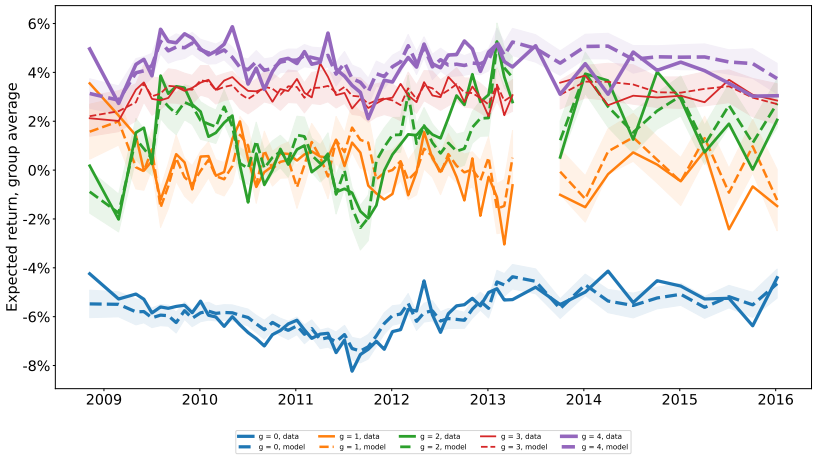


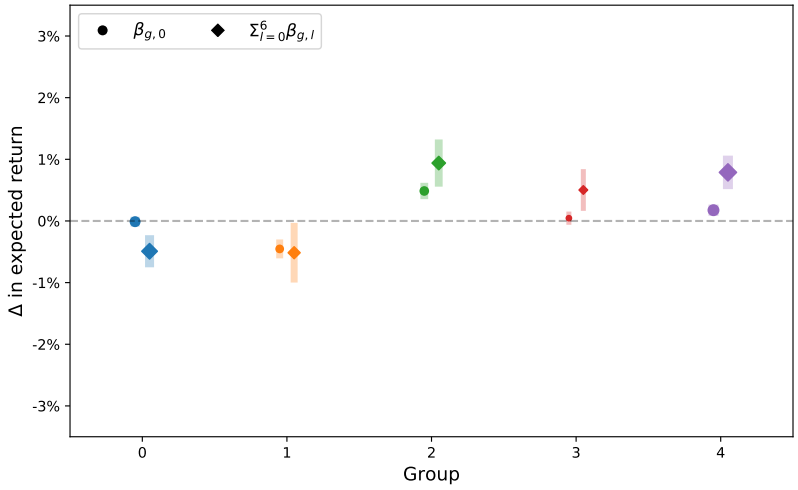
Figure C.7: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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. Where within survey and group means consist of less than 15 observations, we do not plot the series, resulting in a gap. Some ALP surveys had a smaller number of individuals taking part.

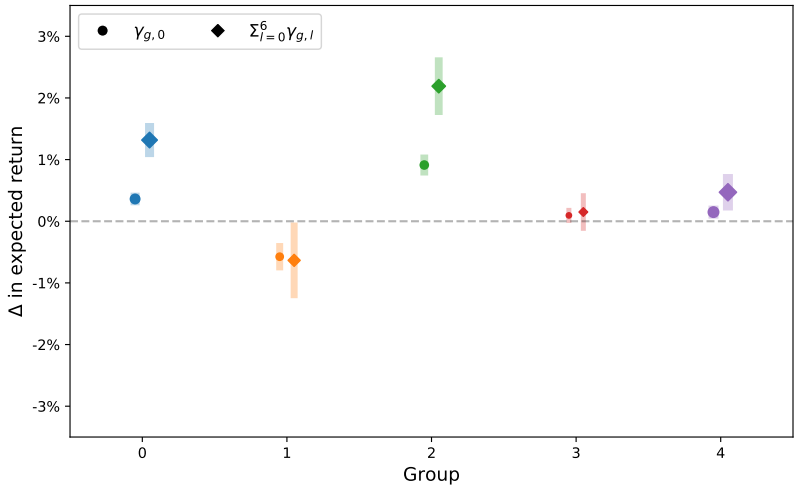
Figure C.8: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



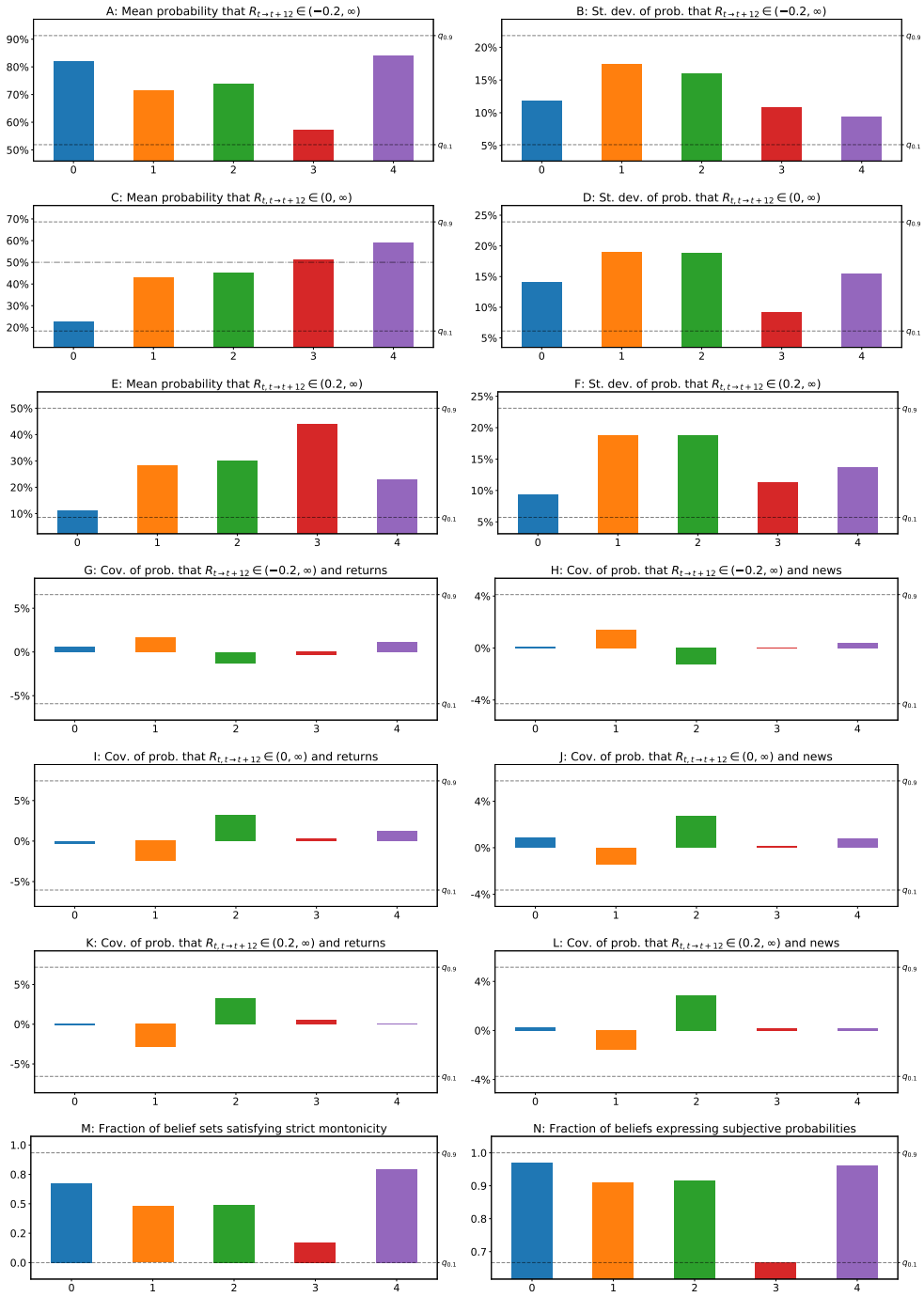
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



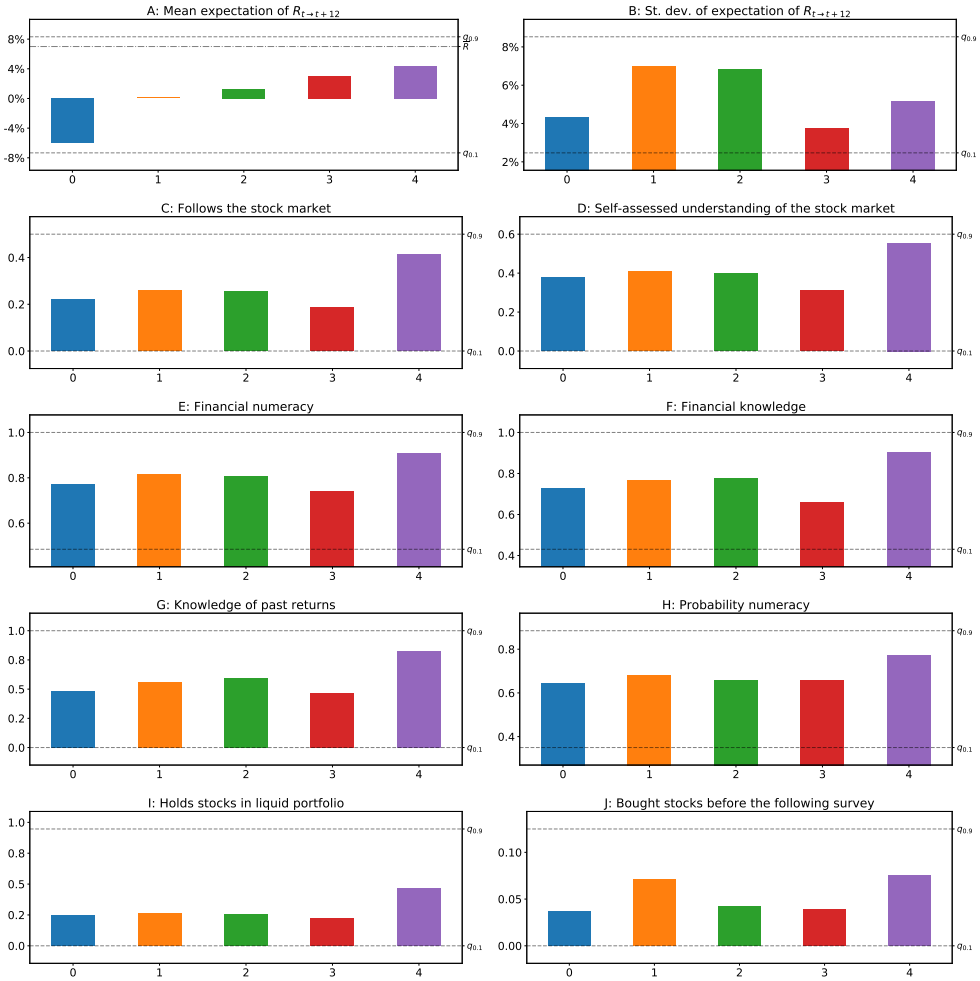
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.9: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.10: Characteristics not used for classification by unobserved heterogeneity group



N = 2042, 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).

Table C.5: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2	3	4
Pessimists	0.83	0.04	0.07	0.00	0.06
Mean Reverters	0.01	0.88	0.08	0.04	0.00
Extrapolators	0.00	0.00	1.00	0.00	0.00
Ignorants	0.00	0.00	0.00	1.00	0.00
Sophisticates	0.00	0.07	0.12	0.02	0.80

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

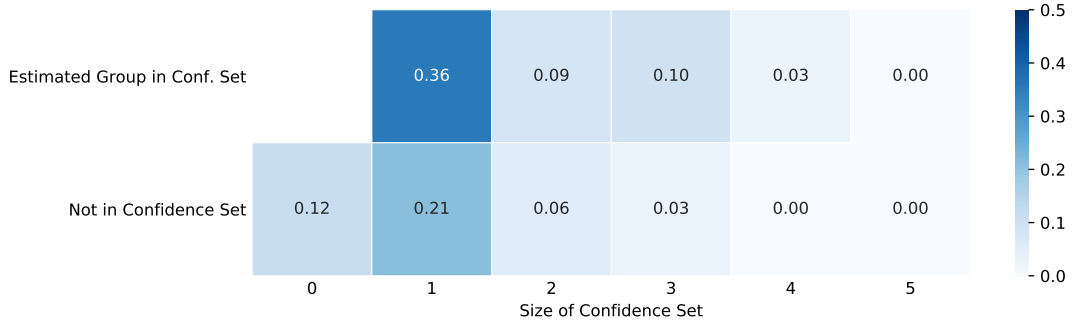
C.4 $\Pr(R_{t\to t+12} > 0)$ as the dependent variable

This section shows the results when we replace our dependent variable from the main analysis, $E[R_{t\to t+12}]_{i,t}$ with $\Pr(R_{t\to t+12} > 0)_{i,t}$. This substantially reduces the amount of information we use on individual beliefs, but is robust to monotonicity violations that arise when we approximate expectations using all three subjective probabilities.

Table C.6: Group sizes

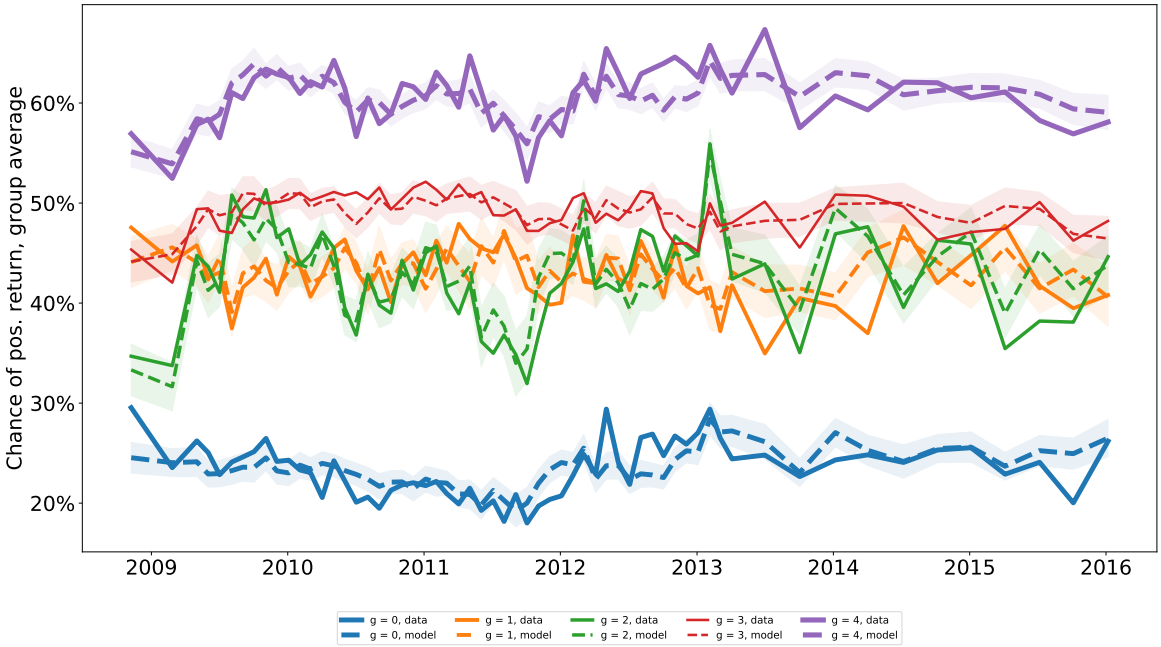
group	N	Share of sample
0	749	0.25
1	584	0.19
2	527	0.17
3	393	0.13
4	777	0.26
Total	3030	

Figure C.11: unit-wise 90% confidence sets by size and inclusion of estimated group



Number in cells refer to its share of the individuals.

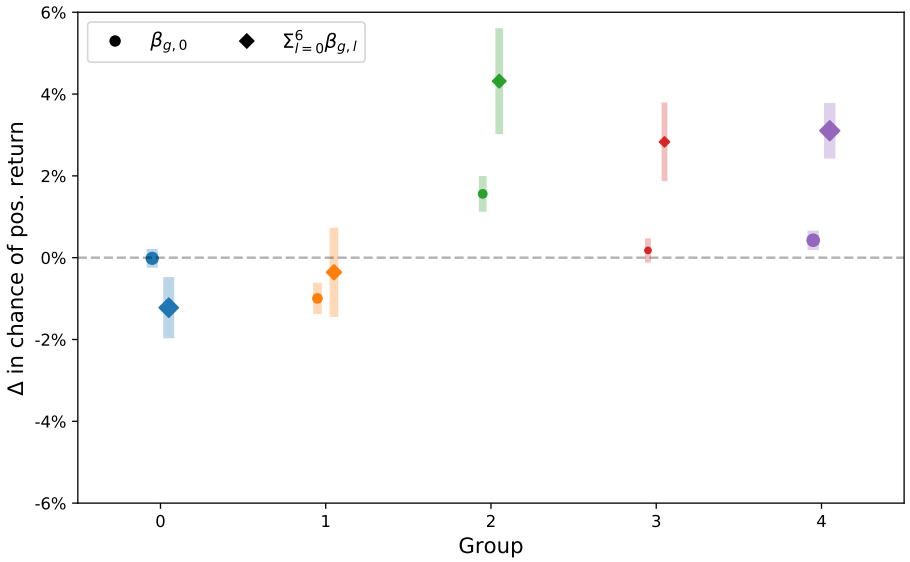
Figure C.12: Data vs. predicted probability that the Dow Jones Index increases, by unobserved group



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.

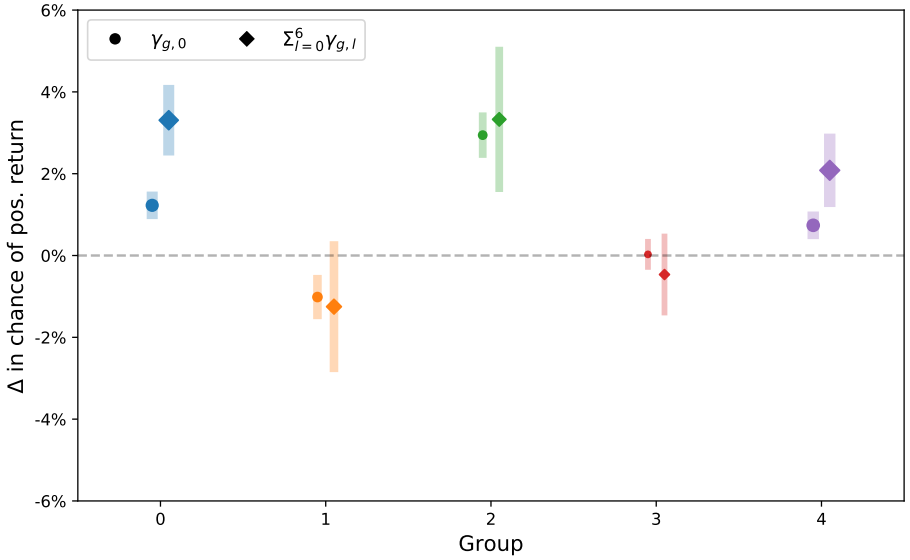
Figure C.13: Effect on the probability that the Dow Jones will go up of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



Dots depict the effect on the probability that the Dow Jones increases 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

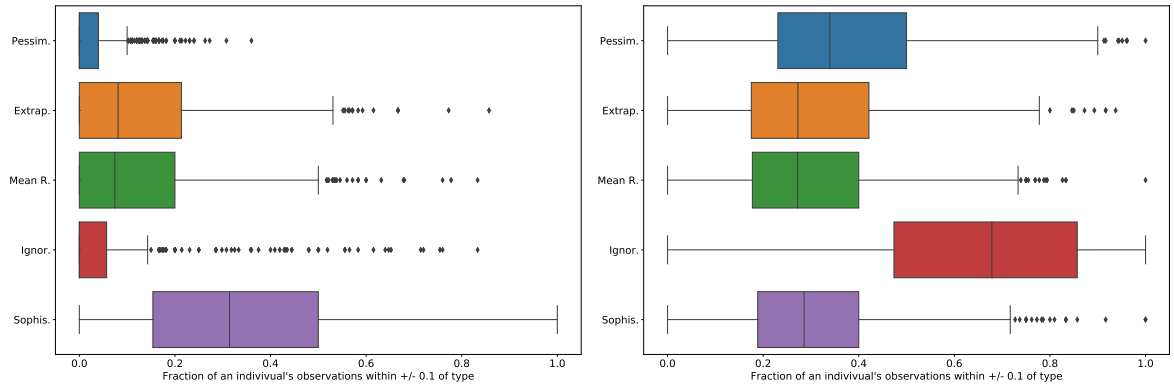


Dots depict the effect on the probability that the Dow Jones increases of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker and line widths proportional to group sizes.

C.5 Comparison with Dominitz-Manski types

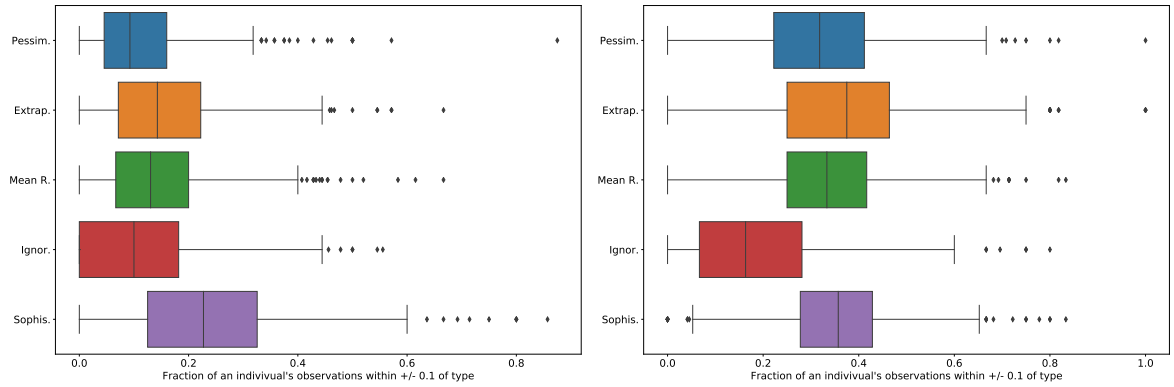
The following plots show how our groups relate to the types of Dominitz and Manski, extended to a much longer panel, by considering the cross-sectional distribution of individual-level fractions of observations close to their specified belief types Random Walk (RW), Persistence (P) and Mean Reversion (MR).

Figure C.14: Cross-sectional distribution of individual compatibility with RW type, by group



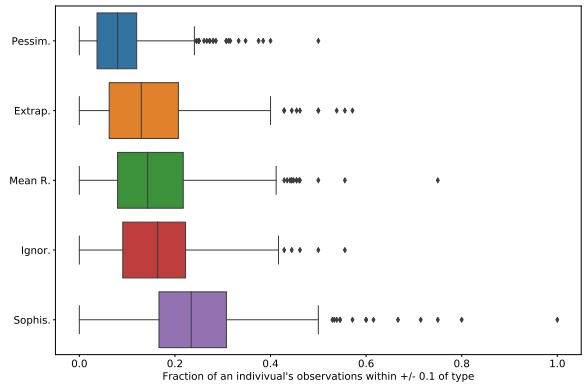
Fully specified belief type | Type def. in terms of response to returns

Figure C.15: Cross-sectional distribution of individual compatibility with P type, by group

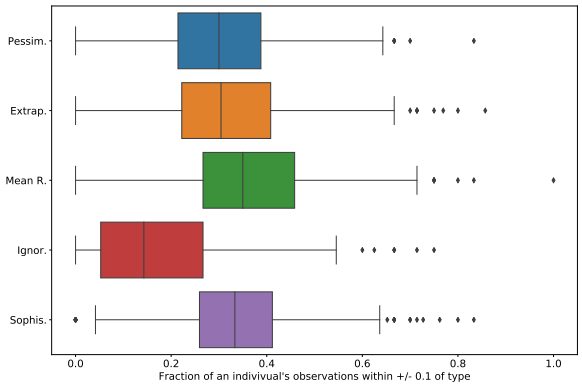


Compatibility with fully specified belief type | Type def. in terms of response to returns

Figure C.16: Cross-sectional distribution of individual compatibility with MR type, by group



Compatibility with fully specified belief type



Type def. in terms of response to returns

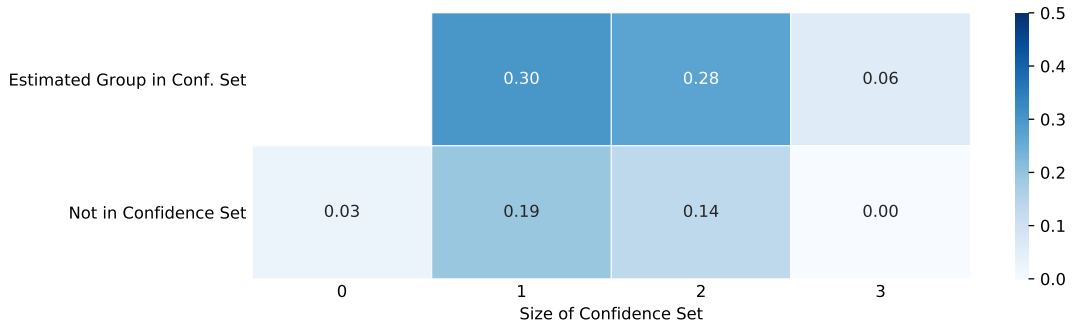
C.6 Three unobserved groups

This section shows the results when we assign individuals to 3 groups.

Table C.7: Group sizes

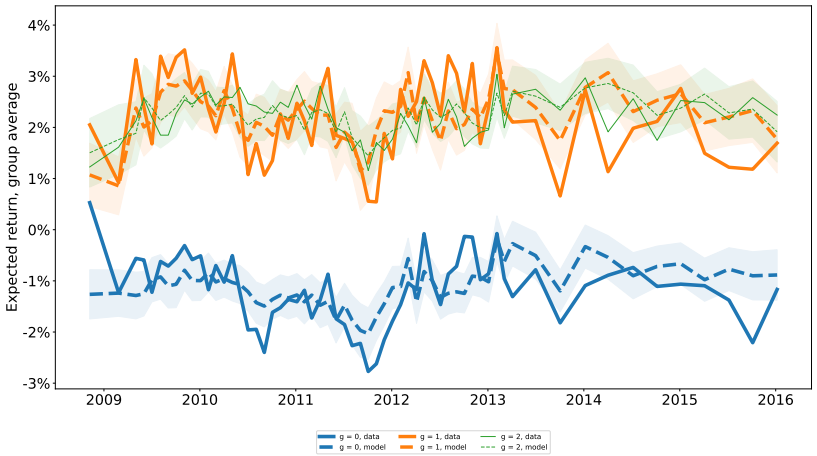
group	N	Share of sample
0	1357	0.45
1	1298	0.43
2	375	0.12
Total	3030	

Figure C.17: unit-wise 90% confidence sets by size and inclusion of estimated group



Number in cells refer to its share of the individuals.

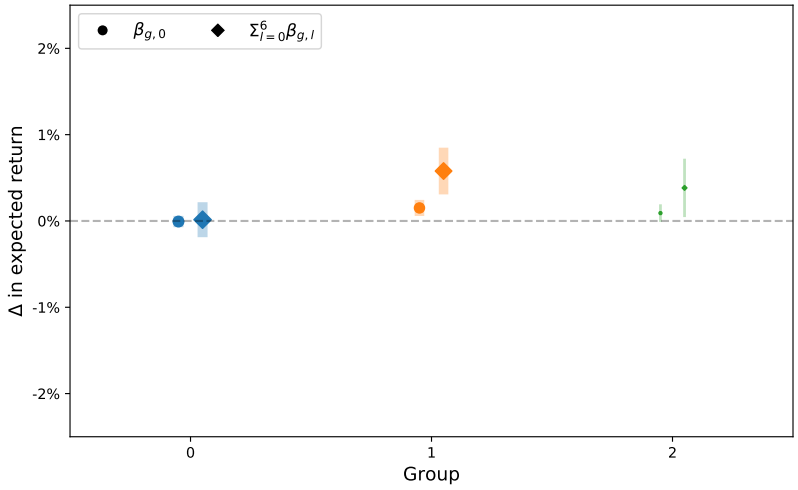
Figure C.18: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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.

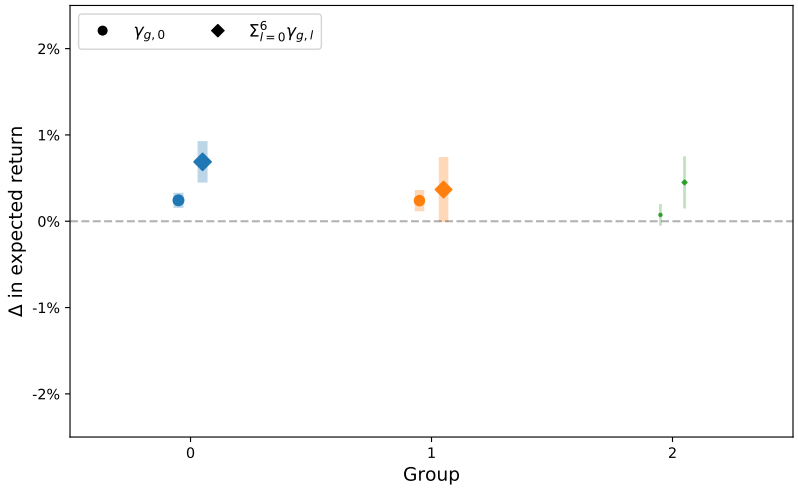
Figure C.19: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



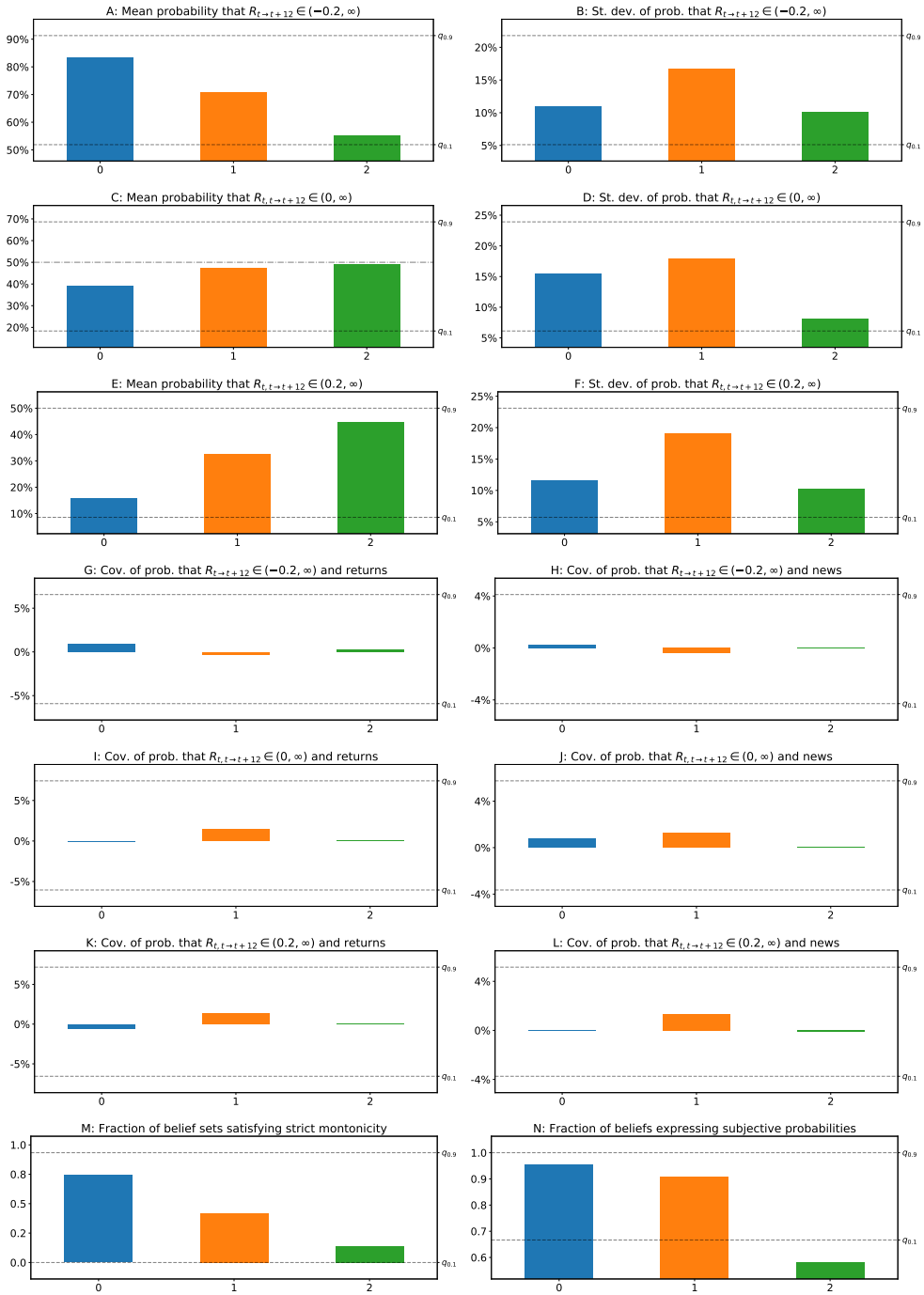
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



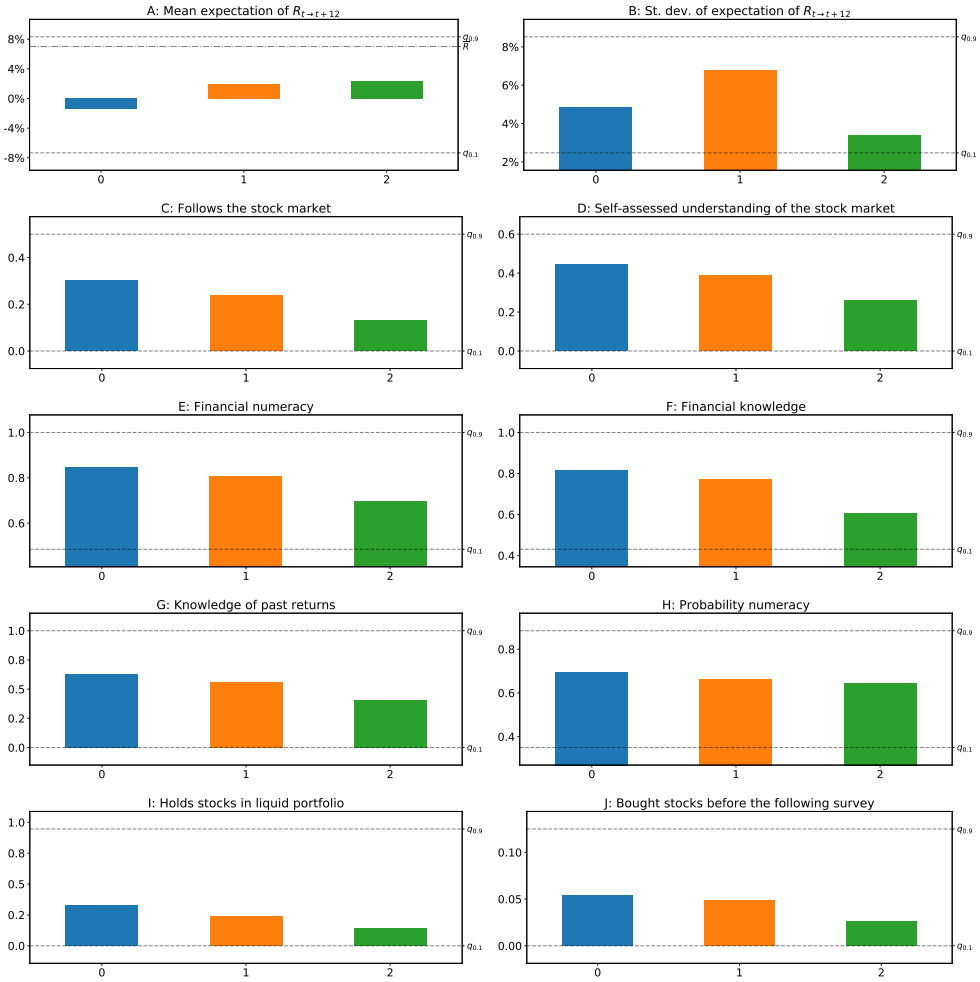
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.20: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.21: Characteristics not used for classification 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).

Table C.8: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2
Pessimists	0.93	0.07	0.00
Mean Reverters	0.15	0.82	0.03
Extrapolators	0.03	0.96	0.01
Ignorants	0.00	0.12	0.88
Sophisticates	0.71	0.28	0.01

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

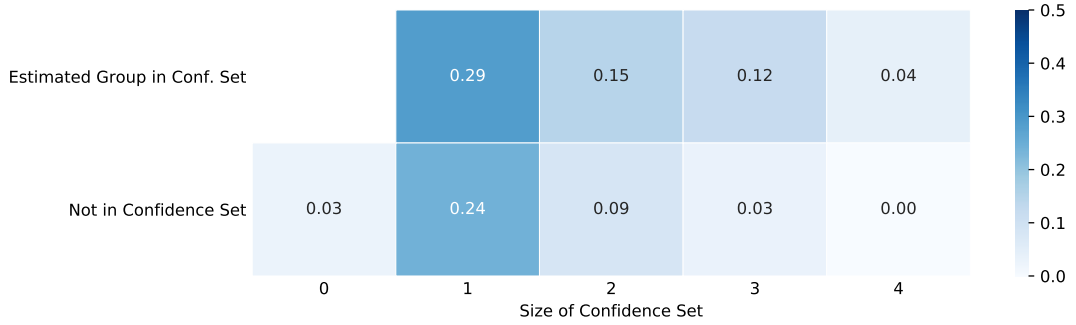
C.7 Four unobserved groups

This section shows the results when we assign individuals to 4 groups.

Table C.9: Group sizes

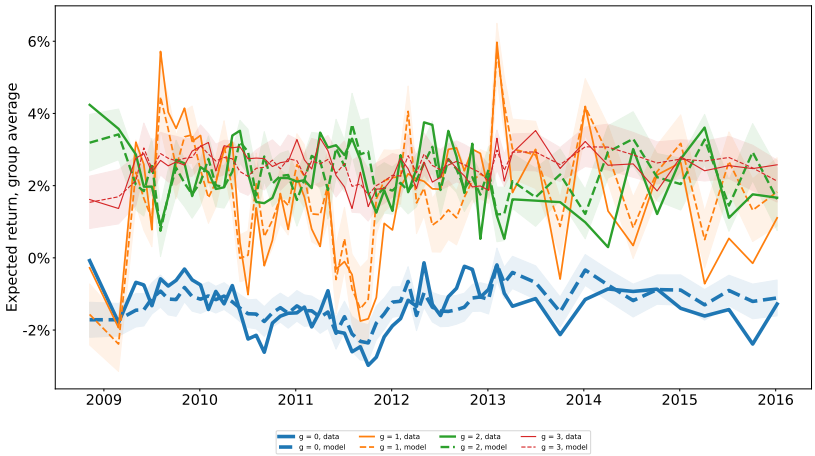
group	N	Share of sample
0	1229	0.41
1	599	0.20
2	810	0.27
3	392	0.13
Total	3030	

Figure C.22: unit-wise 90% confidence sets by size and inclusion of estimated group



Number in cells refer to its share of the individuals.

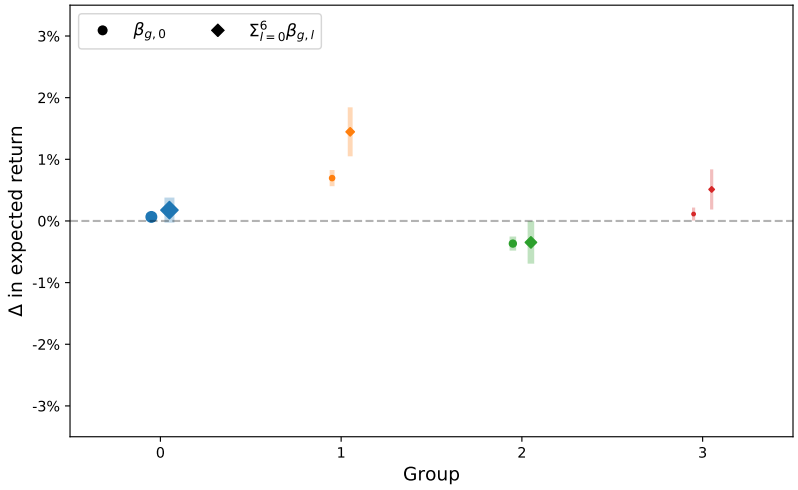
Figure C.23: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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.

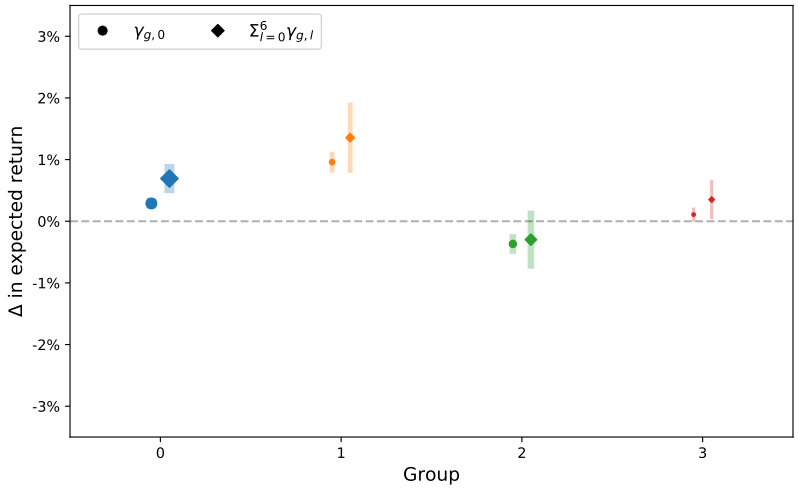
Figure C.24: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



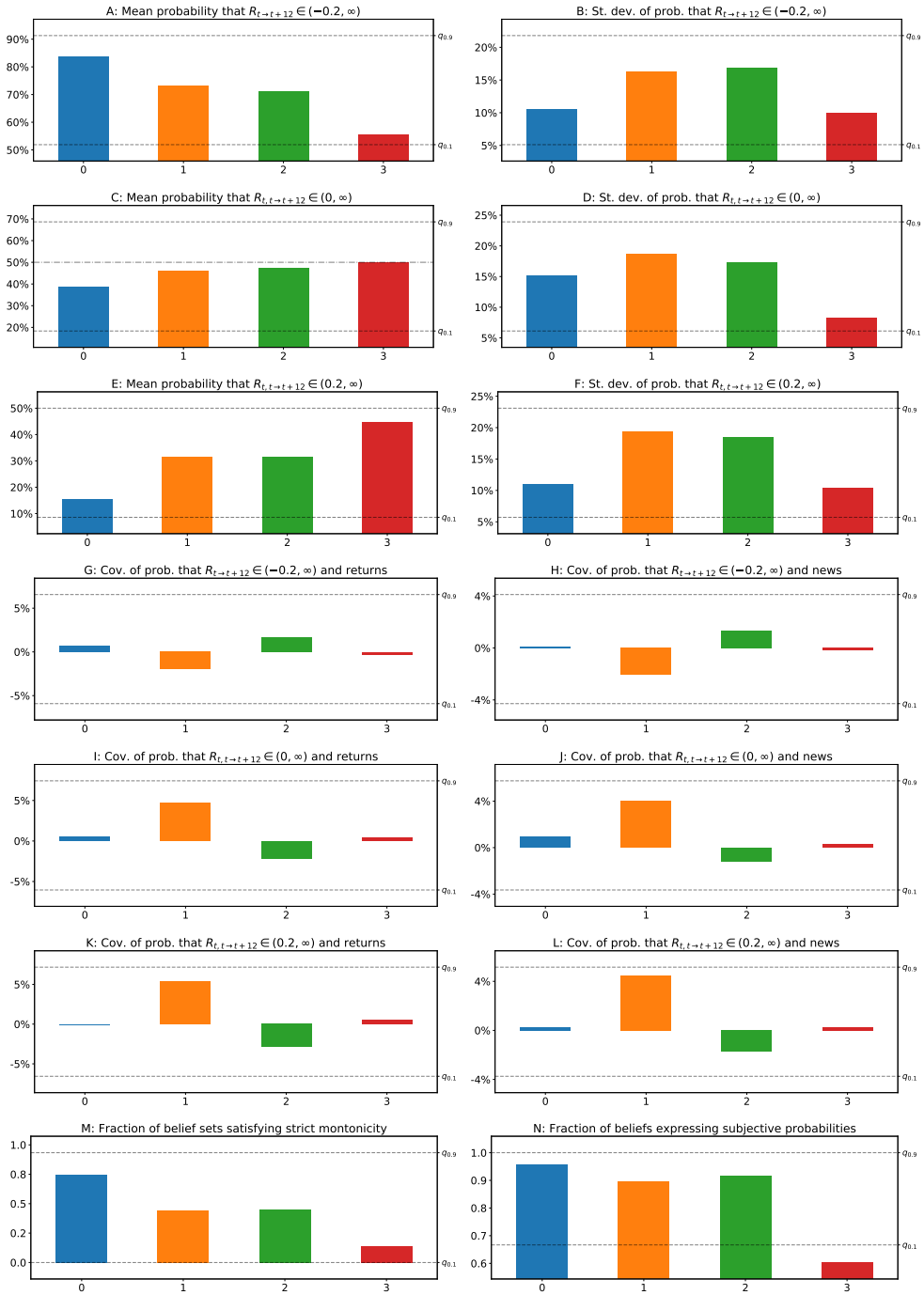
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



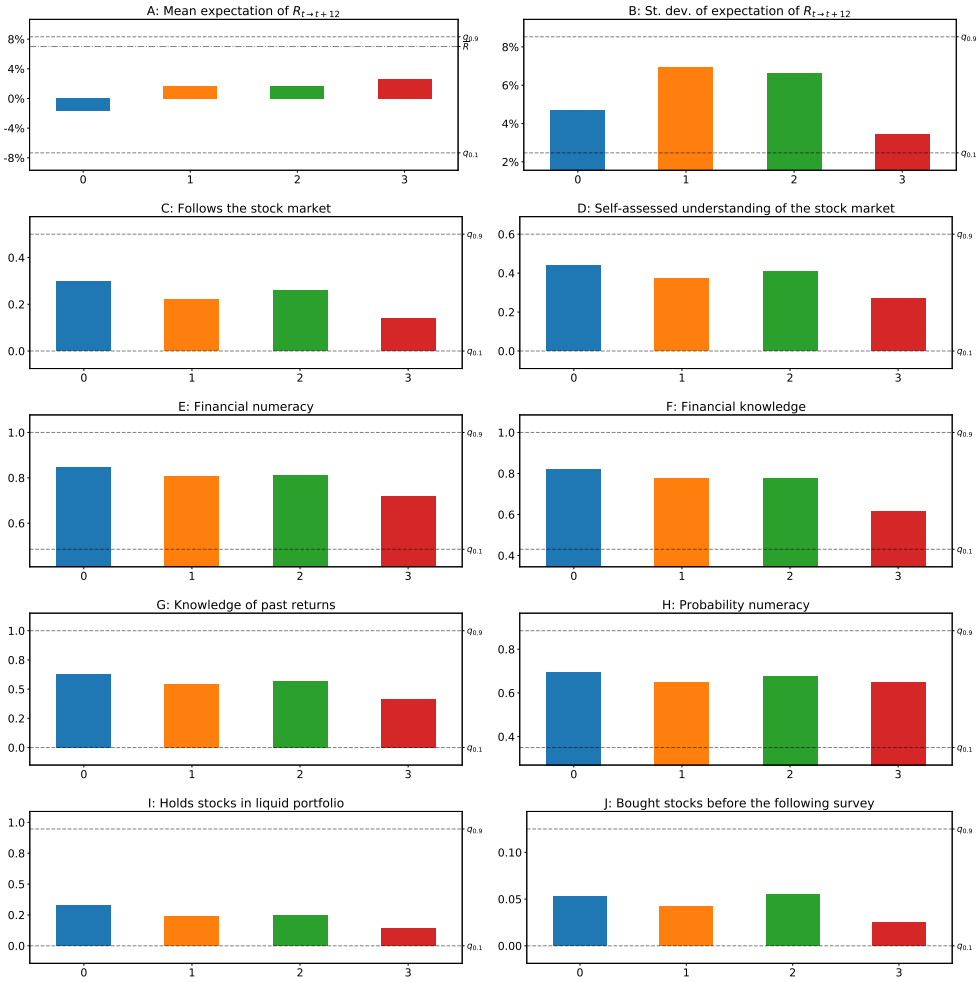
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.25: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.26: Characteristics not used for classification 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).

Table C.10: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2	3
Pessimists	0.94	0.01	0.05	0.00
Mean Reverters	0.01	0.00	0.99	0.00
Extrapolators	0.00	0.97	0.02	0.01
Ignorants	0.00	0.00	0.04	0.96
Sophisticates	0.67	0.10	0.22	0.01

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

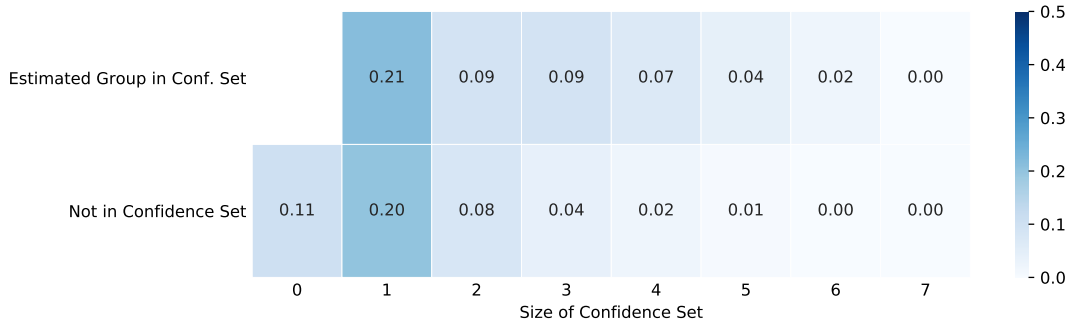
C.8 Seven unobserved groups

This section shows the results when we assign individuals to 7 groups.

Table C.11: Group sizes

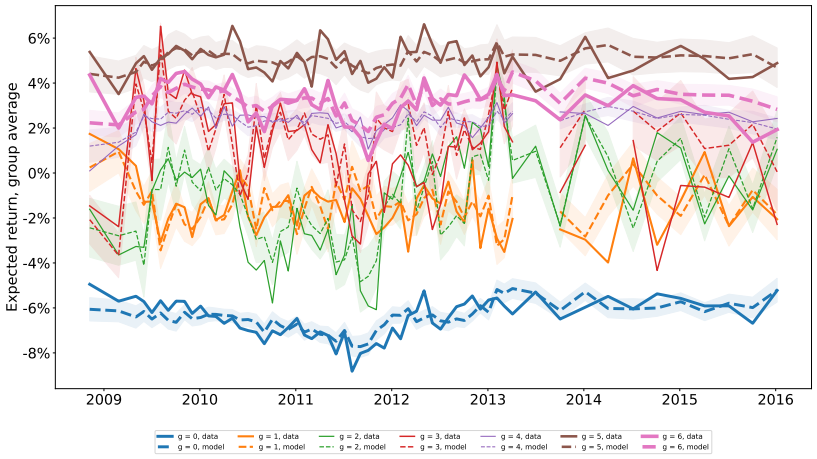
group	N	Share of sample
0	599	0.20
1	424	0.14
2	249	0.08
3	296	0.10
4	216	0.07
5	520	0.17
6	726	0.24
Total	3030	

Figure C.27: unit-wise 90% confidence sets by size and inclusion of estimated group



Number in cells refer to its share of the individuals.

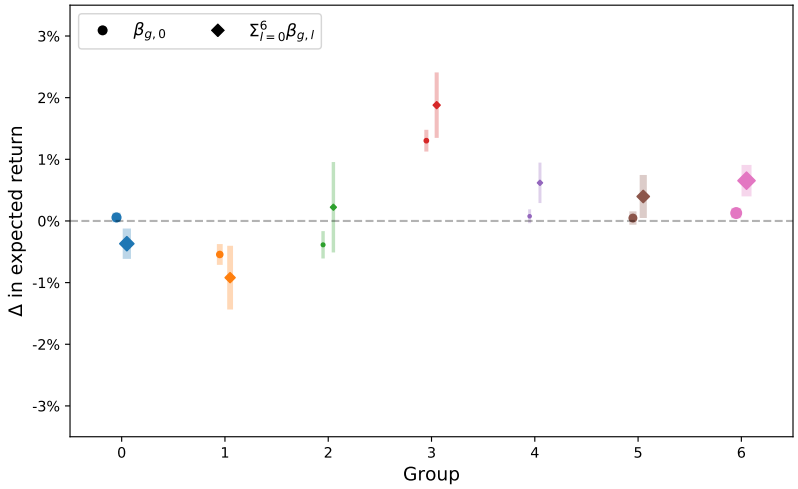
Figure C.28: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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. Where within survey and group means consist of less than 15 observations, we do not plot the series, resulting in a gap. Some ALP surveys had a small number of individuals taking part.

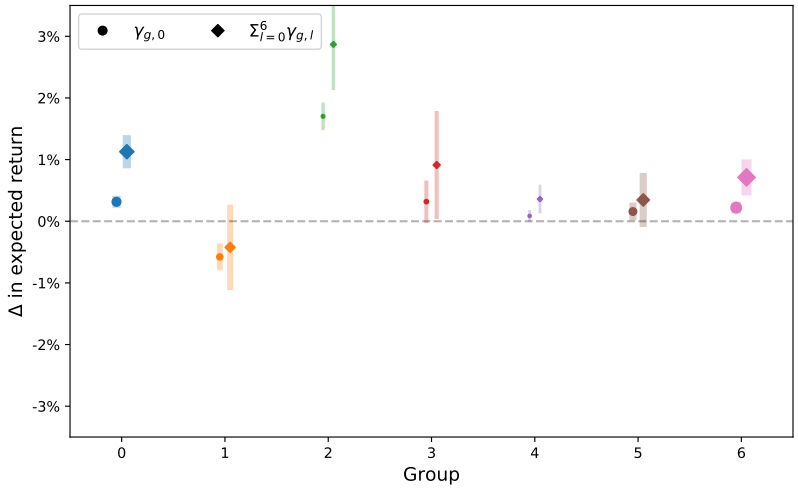
Figure C.29: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



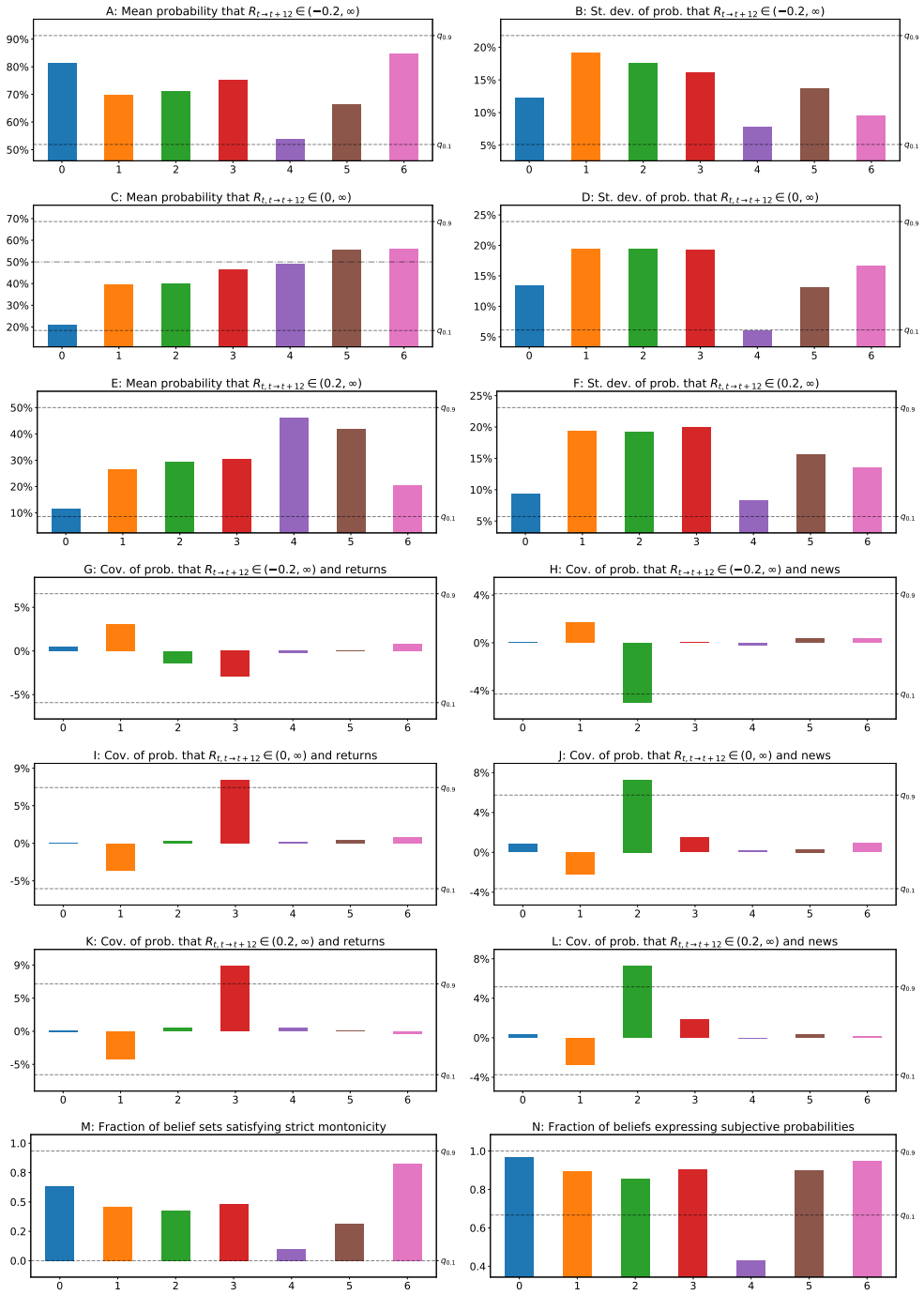
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



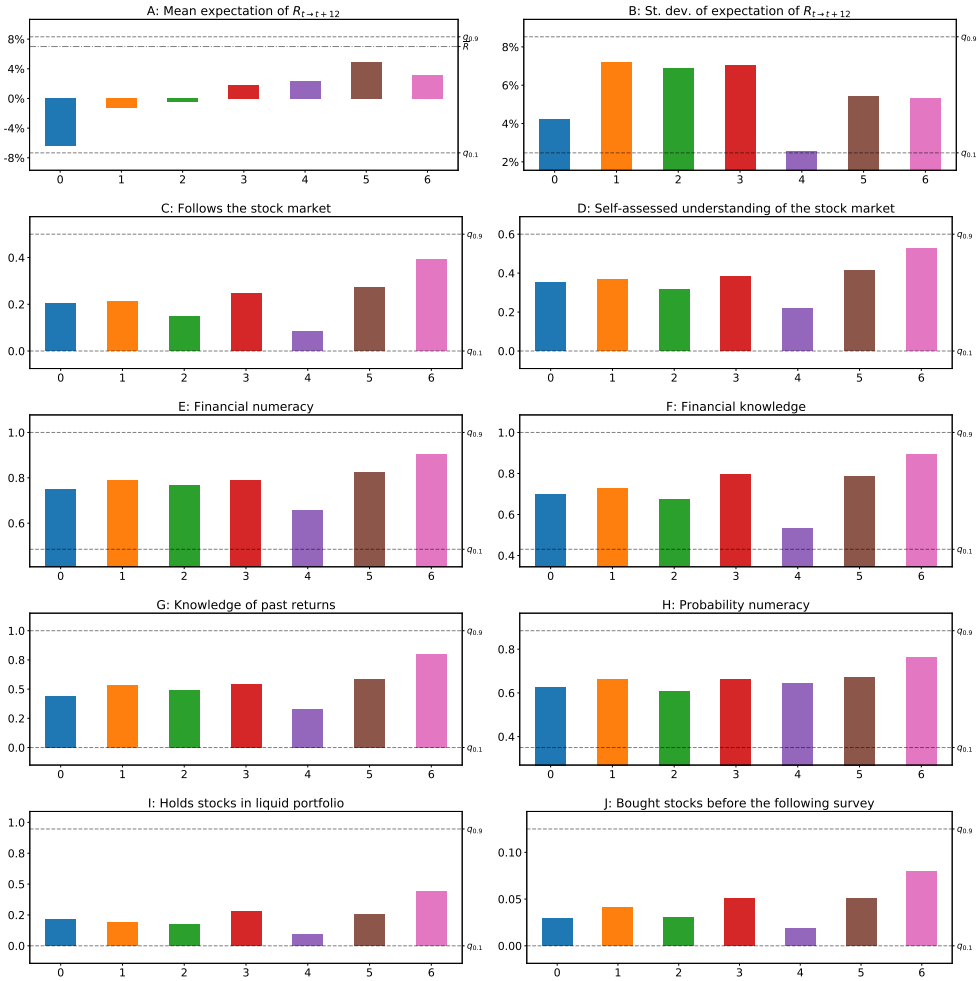
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.30: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.31: Characteristics not used for classification 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).

Table C.12: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2	3	4	5	6
Pessimists	0.79	0.02	0.02	0.02	0.00	0.01	0.15
Mean Reverters	0.00	0.69	0.02	0.02	0.00	0.26	0.01
Extrapolators	0.01	0.01	0.43	0.43	0.00	0.12	0.01
Ignorants	0.00	0.01	0.01	0.01	0.54	0.44	0.00
Sophisticates	0.00	0.00	0.00	0.06	0.00	0.17	0.78

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

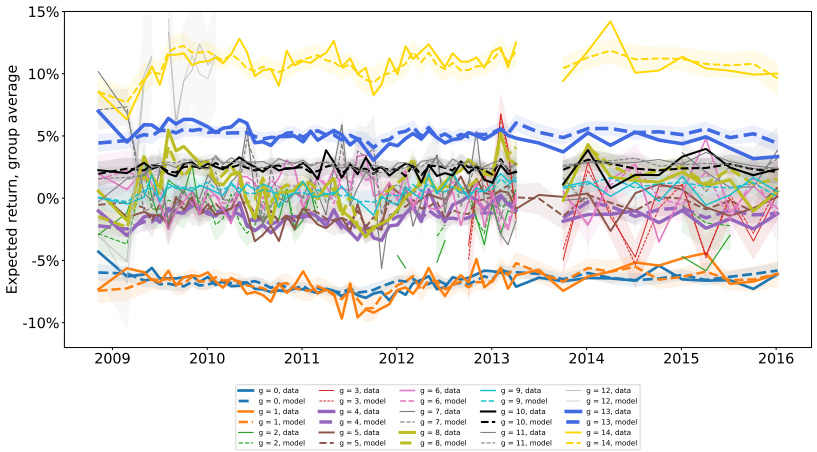
C.9 Fifteen unobserved groups

This section shows the results when we assign individuals to 15 groups.

Table C.13: Group sizes

group	N	Share of sample
0	287	0.09
1	263	0.09
2	124	0.04
3	80	0.03
4	364	0.12
5	207	0.07
6	165	0.05
7	124	0.04
8	322	0.11
9	154	0.05
10	219	0.07
11	113	0.04
12	22	0.01
13	379	0.13
14	207	0.07
Total	3030	

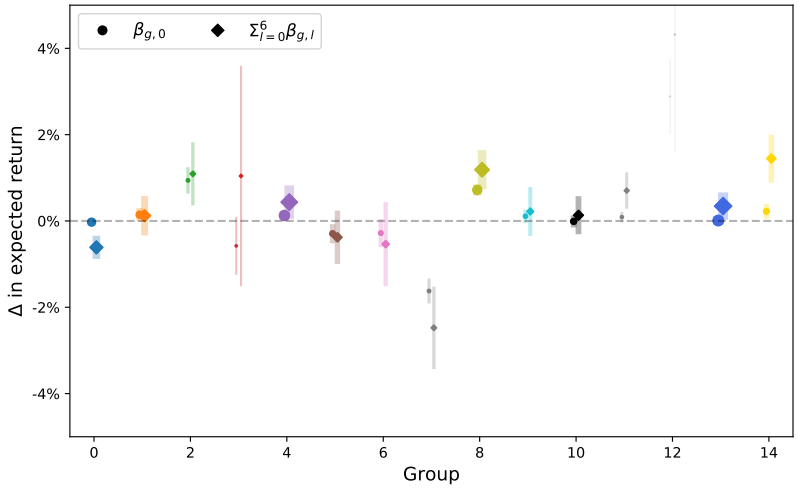
Figure C.32: Data vs. predicted expected return of the Dow Jones index, by unobserved group



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. Where within survey and group means consist of less than 15 observations, we do not plot the series, resulting in a gap. Some ALP surveys had a small number of individuals taking part.

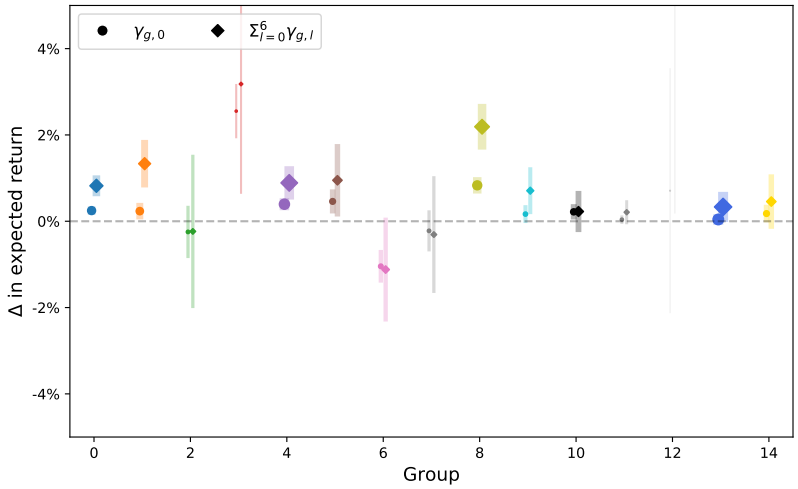
Figure C.33: Effect on expected returns of increases in past returns and tonality of economic news, by group

(a) Effect of past returns



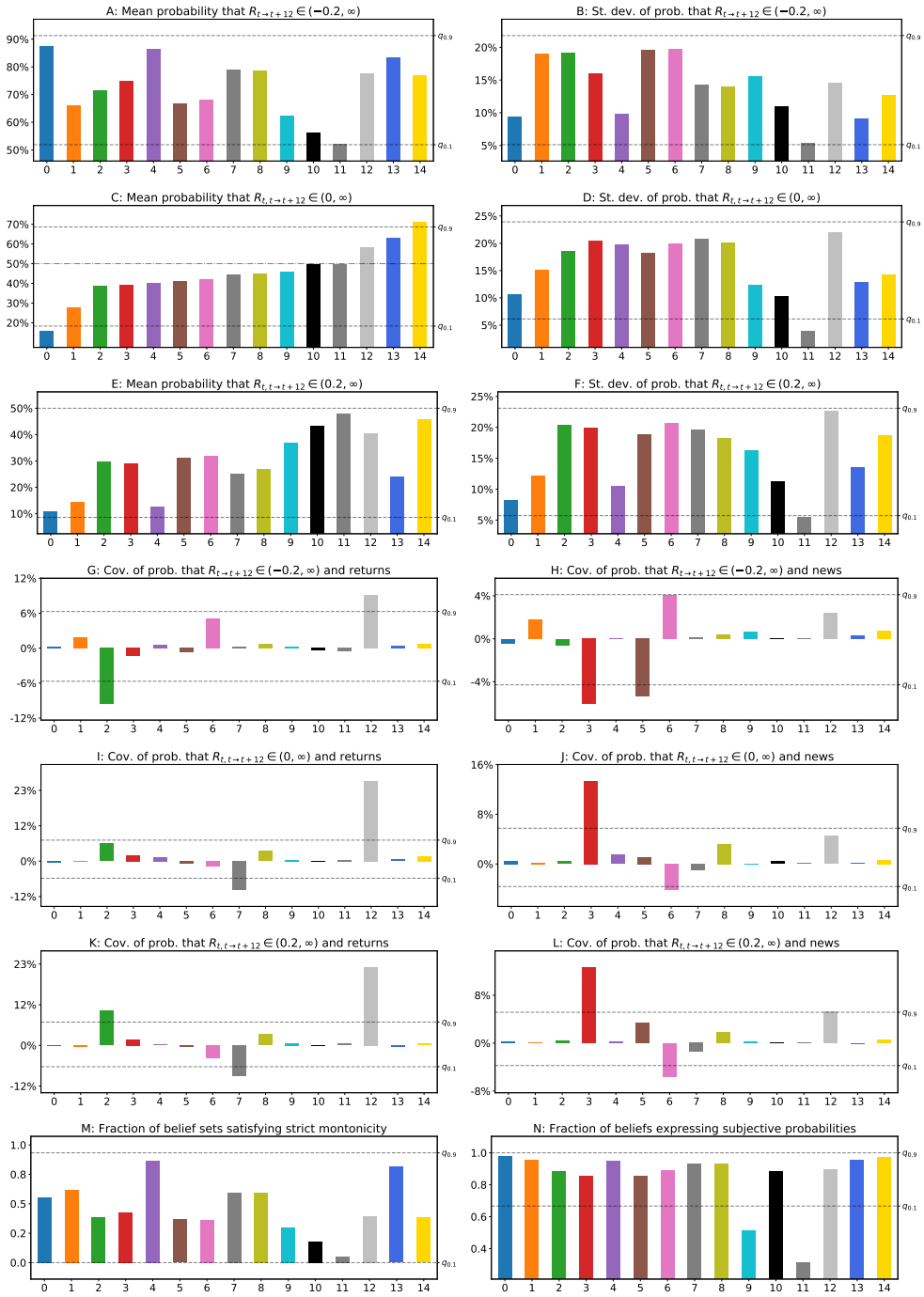
Dots depict the effect on expected returns of a one standard deviation increase in the most recent monthly return of the Dow Jones. Diamonds depict the summed effect in the most recent, plus six preceding monthly returns of the Dow Jones. Shaded lines show the width of 95% confidence intervals. Marker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

(b) Effect of past tonality of economic news



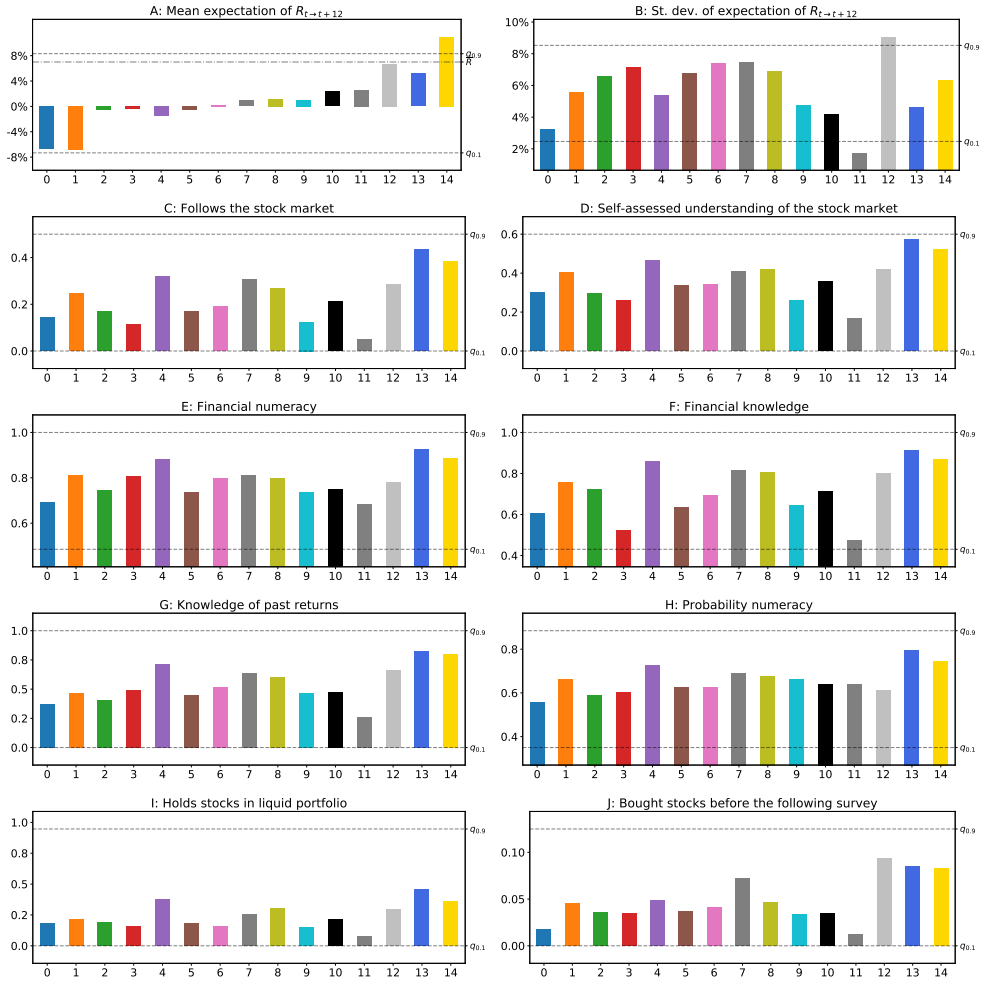
Dots depict the effect on expected returns of a one standard deviation increase in the most recent tonality of economic news over one month. Diamonds depict the summed effect in the most recent, plus six preceding tonalities of economic news. Shaded lines show the width of 95% confidence intervals. Marker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

Figure C.34: Moments used for classification by unobserved heterogeneity group



Bars show the group means of the 14 individual moments used to classify individuals via the k-means algorithm. Dashed lines are the bottom and top decile with respect to the individuals of all groups taken together.

Figure C.35: Characteristics not used for classification 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).

Table C.14: Comparison of estimated groups

Group from alternative specification Group from main specification	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Pessimists	0.38	0.22	0.01	0.00	0.30	0.02	0.00	0.01	0.05	0.00	0.01	0.00	0.00	0.01	0.00
Mean Reverters	0.00	0.14	0.03	0.00	0.00	0.11	0.28	0.16	0.07	0.05	0.06	0.00	0.00	0.00	0.09
Extrapolators	0.00	0.02	0.19	0.15	0.00	0.24	0.00	0.00	0.25	0.03	0.02	0.00	0.04	0.00	0.05
Ignorants	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.26	0.41	0.28	0.00	0.00	0.04
Sophisticates	0.00	0.01	0.00	0.00	0.18	0.00	0.00	0.03	0.14	0.01	0.01	0.00	0.00	0.48	0.15

Each row shows how individuals assigned to a given group in our main specification are allocated across groups for a different specification.

D Additional details on stock ownership and trading

Table D.1: Regression coefficients underlying Figure 7

	Bought stocks in following period
Pessimists	-1.655***
	(0.066)
Mean Reverters	-1.560***
	(0.073)
Extrapolators	-1.768***
	(0.069)
Ignorants	-1.920***
	(0.083)
Sophisticates	-1.410***
	(0.047)
E[R]	0.010***
	(0.002)
N · T	70114
Pseudo R^2	0.02

N = 3029. Probit regression of the dummy indicating whether stocks were bought in the next period (if within 120 days) on expected returns and group indicators. Expected returns in percentage points. Standard errors clustered at individual level.

Table D.2: Stock buying v expected returns

	Average marginal effect	Std. err.
Pessimists	0.09	0.03
Mean Reverters	0.12	0.03
Extrapolators	0.09	0.02
Ignorants	0.07	0.02
Sophisticates	0.16	0.04

N = 3029. From a probit regression of the dummy indicating whether stocks were bought in the next period (if within 120 days) on expected returns and group indicators. Both units in percentage points. Standard errors clustered at individual level. Average marginal effects calculated within group.

E Additional details for forecast error analysis

Table E.1 shows the same analysis as Table 6 in the main text, but uses our entire sample as the basis.

Table E.1: Predictability of forecast errors with forecast revisions, full sample

	Pooled OLS	Pooled OLS w groups
Forecast Revision	-0.51	
	(0.02)	
Forecast Revision, Pessimists		-0.52
		(0.02)
Forecast Revision, Mean Reverters		-0.50
		(0.01)
Forecast Revision, Extrapolators		-0.50
		(0.03)
Forecast Revision, Ignorants		-0.46
		(0.03)
Forecast Revision, Sophisticates		-0.53
		(0.04)
R^2	0.10	0.21
N · T	74165	74165

N = 3030. OLS estimates. Standard errors (clustered by individual and survey) in parentheses.

References

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.

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.

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.

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., 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.

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