Overpersistence Bias in Individual Income Expectations and its Aggregate Implications*

Online Appendix

Filip Rozsypal Kathrin Schlafmann

*Rozsypal: Danmarks Nationalbank and Centre for Macroeconomics (email: firo@nationalbanken.dk); Schlafmann: Copenhagen Business School, Institute for International Economic Studies, Danish Finance Institute (DFI) and Centre for Economic Policy Research (CEPR) (email: ksc.fi@cbs.dk).

Contents

A Further Details about the Empirical Analyses

A.1 Sample Selection

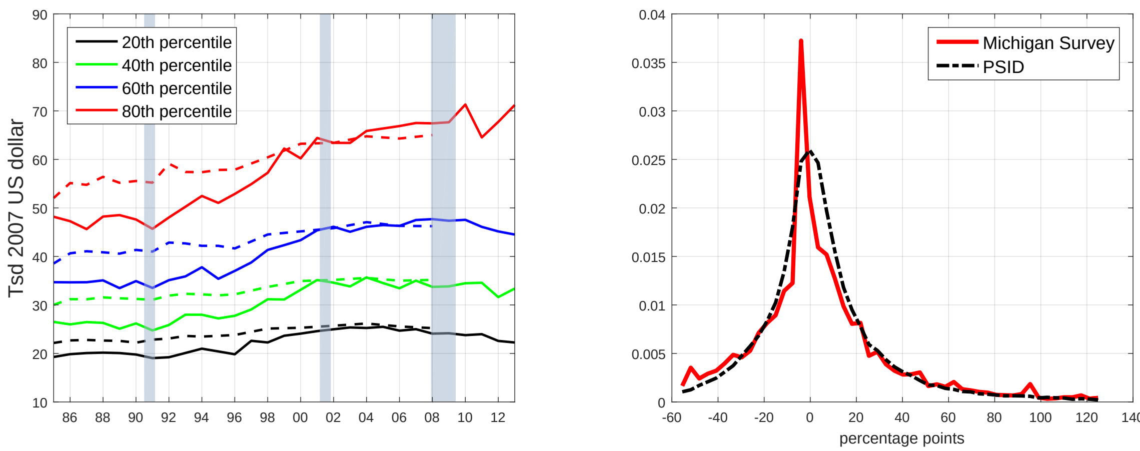
The Michigan Surveys of Consumers interview around 500 households per month of which around one third are re-interviewed after 6 months. The time period that includes precise income information (previously income was only surveyed as bins) is July 1986 - December 2013. Overall, there are observations on 153,241 households (with or without re-interviews). We restrict the sample in the following way: (a) We only select households where the respondent is at most 65 years of age (excludes 30,701 observations). (b) We exclude observations with missing information on demographics (7,605 observations). (c) We exclude observations where the income is lower than the average unemployment benefits in that year (15,525 observations). (d) For households with re-interview we exclude households where the respondent changes between interviews (as identified by the demographics such as gender, age, education, marriage status and racial background, excludes 2,901 observations). Moreover, we exclude households where the number of adults changes between interviews (excludes 3,182 observations). This restriction is made since we are analyzing per adult income in the household, so that changes in the number of adults in the household will reflect changes in this measure of income that might not be anticipated by respondents when they are asked about their income growth expectations.

Overall, this leaves a sample of 88,017 households for which we have full information on demographics as well as inflation expectations (sample INF). 17,500 of these households are both first interviewed in the second half of a year and have a re-interview (sample H2RE). This is the sample for which we have information on realized income growth. Out of sample INF, 41,742 households also provide income expectations and are first interviewed in the first half a year (sample H1), 44,010 provide income expectations and are first interviewed in the second half a year (sample H2).

Figure 1 shows how the income information in our sample compares to the income information in the Panel Study of Income Dynamics (PSID). The PSID is a panel survey that has been running since 1968 which has been widely used to analyze income dynamics. Plot (a) shows that in the first part of the sample real per capita income in the Michigan Surveys is slightly lower than in the PSID. Since the late 1990s, however, the levels of income in both surveys are very similar. Note that we are not using the levels of income in our analysis. Instead, individual income growth rates are the center of our investigation. Plot (b) displays the distribution of these growth rates in the Michigan Surveys and in the PSID. The distribution of income growth is very similar in both surveys. The only difference is that in the Michigan Surveys more households report zero change in nominal income (around 15% of weighted observations, compared 2% in the PSID). To ensure that our results are not driven by these observations, we conduct a robustness check of our main analysis where we exclude all households that report zero income change (see appendix B.2.5). Our results hold and in fact become stronger once these observations are excluded.

Figure 1: Comparison with Income Panel Study of Income Dynamics

(a) percentiles (b) distribution of income growth



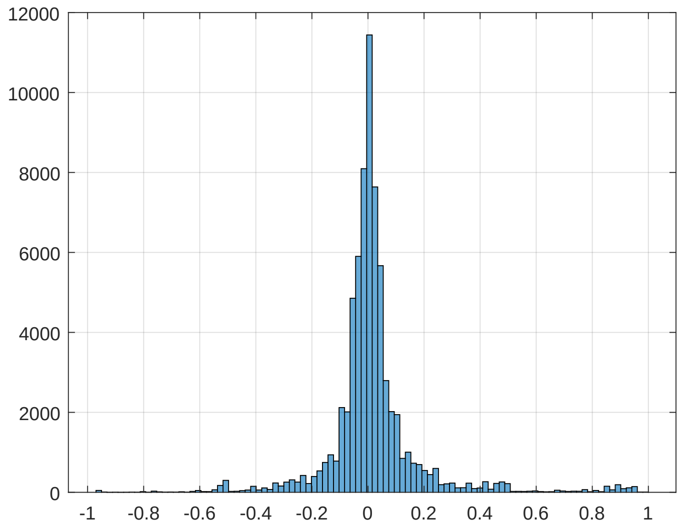
Note: The figure plots a comparison of reported income in the Michigan Surveys and in the Panel Study of Income Dynamics (PSID). Plot (a) shows the percentiles of per capita real income over time: solid lines refer to the Michigan Surveys distribution of income, dashed lines to the corresponding percentiles in the PSID. Plot (b) shows the distribution of real income growth rates in the Michigan Surveys and in the PSID. Since the PSID changed to biannual surveys in 1997, the income growth rates have been constructed from PSID data 1986-1996 only.

Figure 2 displays the distribution of expectations in real income growth and table 1 lists its descriptive statistics. This distribution arguably has some extreme observations. In the main analysis we hence winsorize expectations and forecast errors at the 5% level, i.e. we set the top and bottom 5% of observations to the 95th and 5th percentile of the distribution, respectively. To ensure that our results are not driven by the choice of winsorization threshold we repeat the analysis for varying levels of winsorization (1%, 10% and 25%). The results of this robustness check are shown in appendix B.2.3. All the results remain, merely the magnitudes of the effects become smaller as we remove high and low expectations more and more aggressively.

A.2 Details about the Imputation Procedure

To increase the overlap of expectations and realizations we impute income growth realizations using the information of households with similar household characteristics who report

Figure 2: Distribution of real income expectations



Note: The figure plots the histogram of individual real income growth expectations. For graphical reasons a total of 27 observations (0.04% of all observations) have been removed since they are larger than +100%.

Table 1: Descriptive statistics of real income growth expectations

mean	std	min	p1	p5	median	p95	p99	max
0.018	0.173	-0.967	-0.505	-0.190	0	0.262	0.857	2.1

their income growth for the relevant period. In the example of figure 1 in the main paper, households interviewed for the first time between July 2002 and December 2002 report both their income in 2001 as well as their income in 2002. We can hence use their income in 2001 as well as all available household characteristics to predict their income growth 2001-2002. We then use this relationship to impute income growth 2001-2002 for all households interviewed for the first time in January 2002 to June 2002. The equation that we use to impute income growth realizations is the following:

$$g_{i,t+1} = \alpha + \beta X_{i,t} + \varepsilon_{i,t} \tag{1}$$

where $g_{i,t+1}$ is the growth rate in income of individual i from year t to year $t+1$ and $X_{i,t}$ includes a quadratic term in $\log(income_{i,t})$, a quadratic term in age, as well as indicators for education, gender, ethnic background, marriage status, number of adults, region, income growth expectations, inflation expectations and household weight in the survey. The imputation procedure is implemented as a multiple imputations algorithm using the predictive mean matching method with 5 nearest neighbors and 25 imputations. The imputation procedure is done separately for each survey year, using the observations from sample H2RE which report income changes for the respective year.

Figure 1(c) in the main paper shows that for January households the overlap between expectation and imputed realization is now perfect. For February to June this overlap decreases but is still larger than the maximum overlap we obtain for July to December households on directly reported data. Moreover, for January to June households we do not need any re-interview so that we can use all observations in the data, not only the ones with re-interview. This greatly increases the sample size: We are able to obtain income growth realizations (and thus forecast errors) for the whole sample H1.

Furthermore, we can also increase the overlap for July to December households by imputing income changes for the following year. In the example of figure 1 in the main paper we use the information provided by households interviewed for the first time in July to December 2003 to impute income growth 2002-2003 for the households first interviewed in July to December 2002. This increases the overlap between their expectations and imputed realizations. The largest overlap is 11 months for December households, which is close to perfect. Note that for this step we base the imputation on the income that households reported in their second interview. Unlike in the case of the sample H1, we are hence only able to impute income changes for households who have a re-interview. Combined with the imputed sample H1 this generates the main sample of forecast errors of 58,369 observations (sample MAIN). Table 2 shows the distribution of imputed individual income growth rates in this sample compared to the directly reported income growth rates in sample H2RE. The distribution in the imputed data is very close to the distribution of the original data.

Table 2: Distribution of real reported income changes and imputed values

	mean	p5	p25	p50	p75	p95
directly reported	0.034	-0.378	-0.097	-0.015	0.133	0.572
imputed	0.032	-0.365	-0.103	-0.016	0.130	0.577

Note: The table compares the distribution of imputed individual growth rates in real income in sample MAIN with the growth rates in directly reported income in sample H2RE.

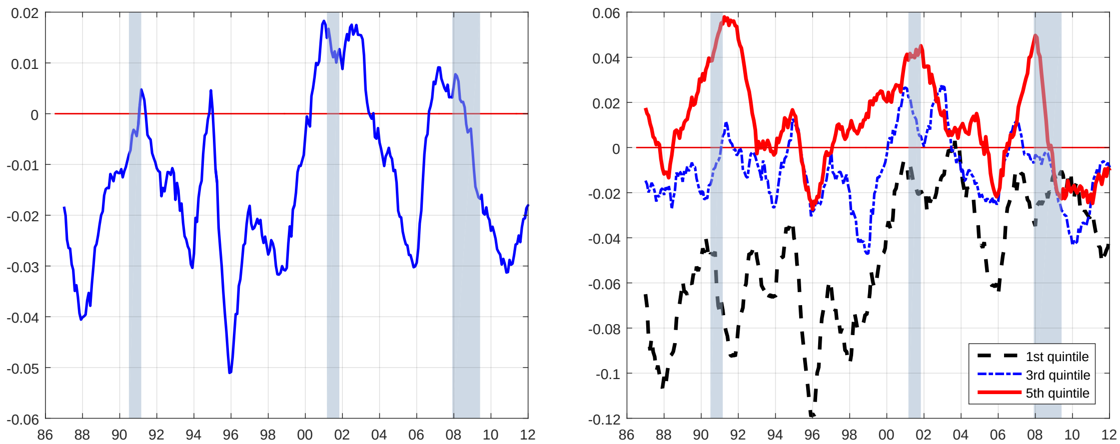
The main analyses reported in this paper are conducted on the sample MAIN where realized income growth has been imputed to maximize both the timing overlap and the number of observations. However, we have conducted robustness checks on the following subsamples: JAN (households with interview in January, income growth imputed, overlap perfect: 6,973 observations); DEC (households with first interview in December, income growth imputed, overlap close to perfect: 2,723 observations); JULY (households with interview in July, directly reported income growth, maximum overlap for directly reported data: 2,805 observations). Whenever imputed income growth is used, standard errors account for the additional uncertainty using multiple imputation procedures and standard errors based on Rubin (1987), Barnard and Rubin (1999) and Reiter (2007).

B Additional Empirical Results & Robustness Checks

B.1 Time Series Plots of Errors in Nominal Income

Figure 3: Expectation errors in nominal income growth

(a) mean (b) mean by income quintile



Note: The figure plots the 12-month moving average of mean expectation errors in individual nominal income growth. Expectation errors are winsorized at 5% and 95%. Data from the Michigan Surveys of Consumers and own calculations. Grey areas represent NBER recessions. On the y-axis, 0.01 corresponds to 1 percentage point.

B.2 Robustness Checks

This appendix contains robustness checks to the specification in the main text. The first robustness check is to include interaction terms of income quintiles with age bins and education dummies. Most of these interaction terms are not significant and the relationship between expectation errors and income quintiles is robust to this change: it remains statistically and economically significant and of very similar magnitude as in the main specification. In a second robustness check we control for cohort effects, in one specification instead of age and in another specification instead of time effects (and include dummies for month of the interview to control for seasonal effects). Our results are virtually unchanged by these alternative controls. The third robustness check is to limit our analysis only to the period 2000 and later. Our results are qualitatively the same as in the main specification. The magnitudes of the effects are smaller but still economically and statistically significant. The fourth robustness check varies the thresholds for winsorization to exclude that our findings are driven by outliers. All results remain for all levels of winsorization, merely the magnitudes become smaller as we remove high and low expectations more and more aggressively. The last robustness check excludes observations with zero reported income change (as the fraction of those households seems slightly inflated compared to the PSID). Also here we find that all our results hold so that the households with zero reported income change are not driving our results.

B.2.1 Interaction with Age and Education

Table 3: OLS of forecast error on observables, interaction with education and age

	real	real	nominal	nominal
1st	−0.051 (0.007)	−0.057 (0.010)	−0.047 (0.007)	−0.054 (0.010)
2nd	−0.017 (0.006)	−0.021 (0.010)	−0.016 (0.006)	−0.018 (0.010)
4th	0.019 (0.005)	0.027 (0.009)	0.017 (0.005)	0.025 (0.009)
5th	0.035 (0.006)	0.047 (0.010)	0.032 (0.006)	0.043 (0.011)
no high school	0.013 (0.014)	0.023 (0.027)	0.019 (0.014)	0.030 (0.028)
college	−0.014 (0.004)	−0.008 (0.008)	−0.017 (0.004)	−0.010 (0.008)
age $<$ 35	0.026 (0.005)	0.021 (0.010)	0.026 (0.005)	0.021 (0.010)
$50 \leq$ age < 65	−0.013 (0.004)	−0.015 (0.009)	−0.014 (0.004)	−0.015 (0.009)
1st × no high school		−0.019 (0.030)		−0.021 (0.030)
2nd × no high school		−0.008 (0.034)		−0.011 (0.035)
4th × no high school		0.015 (0.037)		0.013 (0.038)
5th × no high school		0.020 (0.045)		0.021 (0.046)
1st × college		0.005 (0.013)		0.003 (0.013)
2nd × college		0.001 (0.012)		−0.000 (0.013)
4th × college		−0.013 (0.011)		−0.011 (0.011)
5th × college		−0.021 (0.012)		−0.021 (0.012)
1st × age $<$ 35		0.012 (0.015)		0.014 (0.015)
2nd × age $<$ 35		0.007 (0.014)		0.007 (0.014)
4th × age $<$ 35		−0.004 (0.012)		−0.005 (0.012)
5th × age $<$ 35		0.007 (0.013)		0.008 (0.014)
1st × $50 \leq$ age < 65		0.010 (0.015)		0.010 (0.015)
2nd × $50 \leq$ age < 65		0.005 (0.014)		0.003 (0.014)
4th × $50 \leq$ age < 65		−0.003 (0.012)		−0.004 (0.012)
5th × $50 \leq$ age < 65		−0.001 (0.012)		−0.001 (0.012)
Month dummies	57498	57498	57498	57498

Note: The table shows the results from the multiple imputations OLS regression of equation (2) in the main paper (dependent variable is error in either real or nominal income growth on the household level) with additional interaction terms of income quintiles with education and age groups. Additional regressors (coefficients not shown) are a constant, racial background, number of adults in the household, gender, marriage status as well as region and month dummies. Standard errors reported in parentheses; they take the uncertainty induced by the imputation procedure into account.

B.2.2 Controlling for Cohort Effects

Table 4: OLS of expectation errors on household characteristics, controlling for cohort and time effects

	(1) real	(2) real	(3) real	(4) real	(5) nominal	(6) inflation
Income Quintile						
1 (low)	−0.051	−0.046	−0.047	−0.077	−0.048	0.004
	(0.006)	(0.018)	(0.027)	(0.021)	(0.007)	(0.000)
2	−0.017	−0.013	−0.024	−0.039	−0.016	0.002
	(0.006)	(0.017)	(0.024)	(0.020)	(0.006)	(0.000)
4	0.019	0.026	0.028	0.021	0.017	−0.002
	(0.005)	(0.013)	(0.024)	(0.016)	(0.005)	(0.000)
5 (high)	0.034	0.045	0.039	0.064	0.031	−0.004
	(0.006)	(0.015)	(0.022)	(0.017)	(0.006)	(0.000)
Education						
no high school	0.013	0.013	0.018	−0.002	0.019	0.002
	(0.013)	(0.029)	(0.060)	(0.036)	(0.014)	(0.001)
college	−0.014	−0.024	−0.007	−0.036	−0.017	−0.002
	(0.004)	(0.012)	(0.016)	(0.013)	(0.004)	(0.000)
Racial background						
black	0.019	0.024	0.007	0.021	0.023	0.002
	(0.008)	(0.018)	(0.033)	(0.022)	(0.008)	(0.000)
hispanic	0.012	0.005	0.017	0.017	0.017	0.003
	(0.009)	(0.027)	(0.046)	(0.033)	(0.009)	(0.001)
Number of adults						
1	−0.025	−0.003	−0.036	0.019	−0.025	0.001
	(0.009)	(0.026)	(0.039)	(0.042)	(0.010)	(0.001)
3 or more	0.018	0.012	0.017	0.021	0.016	−0.002
	(0.007)	(0.018)	(0.029)	(0.022)	(0.007)	(0.000)
Other family characteristics						
female	−0.008	−0.005	−0.007	−0.008	−0.002	0.005
	(0.004)	(0.010)	(0.016)	(0.012)	(0.004)	(0.000)
not married	0.023	0.003	0.033	−0.011	0.024	0.000
	(0.009)	(0.024)	(0.035)	(0.041)	(0.009)	(0.000)
Region						
North Central	−0.022	−0.023	−0.031	−0.021	−0.023	−0.000
	(0.006)	(0.016)	(0.024)	(0.017)	(0.006)	(0.000)
Northeast	−0.020	−0.021	−0.037	−0.005	−0.020	0.001
	(0.006)	(0.017)	(0.027)	(0.018)	(0.006)	(0.000)
South	−0.018	−0.014	−0.029	0.013	−0.017	0.001
	(0.006)	(0.016)	(0.024)	(0.016)	(0.006)	(0.000)
Constant	0.010	0.013	−0.051	0.114	0.000	−0.017
	(0.052)	(0.084)	(0.111)	(0.093)	(0.054)	(0.002)
Sample	MAIN	JAN	DEC	JULY	MAIN	INF
Observations	58369	6973	2723	2805	58369	88017

Note: The table shows results from the multiple imputations OLS regression of equation (2) in the main paper, where the dependent variable is the household expectation error in real income (columns 1-4), in nominal income (column 5) and in inflation (columns 6). The regressions include cohort dummies and month dummies as additional control. Standard errors reported in parentheses; they take the uncertainty induced by the imputation procedure into account whenever imputed data is used; without imputed data heteroskedasticity-robust standard errors are computed.

Table 5: OLS of expectation errors on household characteristics, controlling for age and cohort effects

	(1) real	(2) real	(3) real	(4) real	(5) nominal	(6) inflation
Income Quintile						
1 (low)	−0.052	−0.047	−0.049	−0.078	−0.048	0.004
	(0.006)	(0.018)	(0.027)	(0.021)	(0.007)	(0.000)
2	−0.017	−0.011	−0.024	−0.038	−0.015	0.002
	(0.006)	(0.017)	(0.024)	(0.020)	(0.006)	(0.000)
4	0.019	0.026	0.029	0.022	0.018	−0.002
	(0.005)	(0.013)	(0.024)	(0.016)	(0.005)	(0.000)
5 (high)	0.035	0.045	0.041	0.065	0.032	−0.004
	(0.006)	(0.015)	(0.022)	(0.017)	(0.006)	(0.000)
Education						
no high school	0.014	0.013	0.021	0.008	0.020	0.002
	(0.013)	(0.029)	(0.059)	(0.036)	(0.014)	(0.001)
college	−0.014	−0.024	−0.007	−0.033	−0.017	−0.002
	(0.004)	(0.012)	(0.016)	(0.013)	(0.004)	(0.000)
Age						
age	−0.003	−0.000	−0.006	−0.006	−0.003	0.001
	(0.002)	(0.004)	(0.007)	(0.005)	(0.002)	(0.000)
age × age	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Racial background						
black	0.019	0.024	0.008	0.019	0.023	0.003
	(0.008)	(0.018)	(0.033)	(0.022)	(0.008)	(0.001)
hispanic	0.013	0.006	0.017	0.014	0.018	0.003
	(0.009)	(0.027)	(0.046)	(0.034)	(0.009)	(0.001)
Number of adults						
1	−0.025	−0.002	−0.037	0.024	−0.024	0.002
	(0.009)	(0.026)	(0.040)	(0.043)	(0.010)	(0.001)
3 or more	0.019	0.012	0.018	0.027	0.017	−0.002
	(0.007)	(0.018)	(0.030)	(0.022)	(0.007)	(0.000)
Other family characteristics						
female	−0.008	−0.006	−0.006	−0.010	−0.003	0.005
	(0.004)	(0.010)	(0.016)	(0.012)	(0.004)	(0.000)
not married	0.023	0.002	0.032	−0.017	0.023	−0.000
	(0.009)	(0.024)	(0.036)	(0.042)	(0.009)	(0.001)
Region						
North Central	−0.022	−0.023	−0.032	−0.023	−0.023	−0.000
	(0.006)	(0.016)	(0.024)	(0.018)	(0.006)	(0.000)
Northeast	−0.020	−0.021	−0.037	−0.007	−0.021	0.001
	(0.006)	(0.017)	(0.026)	(0.018)	(0.006)	(0.000)
South	−0.018	−0.014	−0.029	0.012	−0.017	0.001
	(0.006)	(0.016)	(0.023)	(0.016)	(0.006)	(0.000)
Constant	0.085	0.020	0.078	0.080	0.039	−0.062
	(0.038)	(0.108)	(0.185)	(0.125)	(0.040)	(0.002)
Sample	MAIN	JAN	DEC	JULY	MAIN	INF
Observations	58369	6973	2723	2805	58369	88017

Note: The table shows results from the multiple imputations OLS regression of equation (2) in the main paper, where the dependent variable is the household expectation error in real income (columns 1-4), in nominal income (column 5) and in inflation (columns 6). The regressions include cohort dummies and indicators for the month of year of the interview as additional controls. Standard errors reported in parentheses; they take the uncertainty induced by the imputation procedure into account whenever imputed data is used; without imputed data heteroskedasticity-robust standard errors are computed.

B.2.3 Subsample year 2000 and later

Table 6: OLS of expectation errors on household characteristics, sample year 2000 and later

	(1) real	(2) real	(3) real	(4) real	(5) nominal	(6) inflation
Income Quintile						
1 (low)	−0.031	−0.026	−0.028	−0.013	−0.026	0.005
	(0.008)	(0.022)	(0.036)	(0.029)	(0.009)	(0.001)
2	−0.010	−0.007	−0.024	−0.010	−0.007	0.002
	(0.007)	(0.022)	(0.033)	(0.026)	(0.008)	(0.001)
4	0.014	0.017	0.017	0.038	0.013	−0.002
	(0.006)	(0.018)	(0.034)	(0.025)	(0.007)	(0.001)
5 (high)	0.025	0.029	0.027	0.072	0.020	−0.005
	(0.008)	(0.024)	(0.033)	(0.025)	(0.008)	(0.001)
Education						
no high school	−0.002	−0.023	0.032	0.015	0.000	0.000
	(0.021)	(0.051)	(0.094)	(0.038)	(0.021)	(0.001)
college	−0.011	−0.021	−0.008	−0.016	−0.014	−0.003
	(0.006)	(0.015)	(0.021)	(0.019)	(0.006)	(0.000)
Age						
age	−0.004	−0.003	−0.007	0.003	−0.004	0.000
	(0.002)	(0.004)	(0.009)	(0.007)	(0.002)	(0.000)
age × age	0.000	0.000	0.000	−0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Racial background						
black	0.025	0.045	0.002	0.028	0.026	0.000
	(0.011)	(0.027)	(0.047)	(0.032)	(0.011)	(0.001)
hispanic	0.030	0.021	0.024	−0.016	0.030	0.000
	(0.011)	(0.035)	(0.061)	(0.046)	(0.012)	(0.001)
Number of adults						
1	−0.005	0.019	−0.038	−0.009	−0.004	0.001
	(0.012)	(0.034)	(0.049)	(0.057)	(0.012)	(0.001)
3 or more	0.017	0.002	0.030	−0.010	0.015	−0.002
	(0.008)	(0.025)	(0.041)	(0.028)	(0.008)	(0.001)
Other family characteristics						
female	−0.009	−0.004	−0.009	−0.002	−0.005	0.005
	(0.005)	(0.013)	(0.020)	(0.016)	(0.005)	(0.000)
not married	0.008	−0.015	0.040	0.015	0.009	0.001
	(0.011)	(0.031)	(0.043)	(0.054)	(0.012)	(0.001)
Region						
North Central	−0.019	−0.018	−0.034	−0.003	−0.018	0.000
	(0.009)	(0.021)	(0.030)	(0.024)	(0.009)	(0.000)
Northeast	−0.011	−0.014	−0.019	0.007	−0.011	0.001
	(0.008)	(0.023)	(0.033)	(0.025)	(0.008)	(0.001)
South	−0.012	−0.013	−0.016	0.013	−0.010	0.002
	(0.008)	(0.021)	(0.031)	(0.022)	(0.008)	(0.000)
Constant	0.120	0.095	0.240	−0.064	0.107	−0.018
	(0.051)	(0.100)	(0.199)	(0.148)	(0.053)	(0.003)
Sample	MAIN	JAN	DEC	JULY	MAIN	INF
Observations	27279	3315	1252	1262	27279	40434

Note: The table shows results of the multiple imputations OLS regression of equation (2) in the main paper, where the dependent variable is the household expectation error in real income (columns 1-4), in nominal income (column 5) and in inflation (columns 6). The regressions include month dummies as additional controls and use only observations for sample of year 2000 and later. Standard errors reported in parentheses; they take the uncertainty induced by the imputation procedure into account whenever imputed data is used; without imputed data heteroskedasticity-robust standard errors are computed.

B.2.4 Winsorization at different thresholds

Table 7: OLS of expectation errors on household characteristics - varying Winsorization thresholds

	(1) real	(2) real	(3) real	(4) real
Income Quintile				
1 (low)	−0.052	−0.070	−0.040	−0.019
	(0.006)	(0.009)	(0.005)	(0.003)
2	−0.018	−0.022	−0.014	−0.007
	(0.006)	(0.008)	(0.004)	(0.002)
4	0.019	0.022	0.016	0.009
	(0.005)	(0.006)	(0.004)	(0.002)
5 (high)	0.035	0.041	0.029	0.016
	(0.006)	(0.008)	(0.005)	(0.002)
Education				
no high school	0.014	0.016	0.010	0.004
	(0.013)	(0.017)	(0.011)	(0.005)
college	−0.014	−0.019	−0.010	−0.004
	(0.004)	(0.005)	(0.003)	(0.002)
Age				
age	−0.004	−0.005	−0.004	−0.002
	(0.001)	(0.002)	(0.001)	(0.001)
age × age	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Racial background				
black	0.019	0.020	0.017	0.009
	(0.008)	(0.010)	(0.006)	(0.003)
hispanic	0.013	0.015	0.010	0.005
	(0.009)	(0.012)	(0.007)	(0.004)
Number of adults				
1	−0.025	−0.029	−0.020	−0.011
	(0.009)	(0.012)	(0.007)	(0.004)
3 or more	0.020	0.018	0.017	0.009
	(0.007)	(0.010)	(0.006)	(0.003)
Other family characteristics				
female	−0.008	−0.008	−0.007	−0.004
	(0.004)	(0.006)	(0.003)	(0.002)
not married	0.023	0.027	0.018	0.009
	(0.009)	(0.012)	(0.007)	(0.004)
Region				
North Central	−0.022	−0.026	−0.017	−0.009
	(0.006)	(0.007)	(0.005)	(0.003)
Northeast	−0.020	−0.025	−0.016	−0.008
	(0.006)	(0.007)	(0.005)	(0.003)
South	−0.018	−0.022	−0.014	−0.007
	(0.006)	(0.008)	(0.005)	(0.002)
Constant	0.136	0.136	0.123	0.078
	(0.052)	(0.071)	(0.041)	(0.020)
Winsorization	WIN 5	WIN 1	WIN 10	WIN 25
Observations	58369	58369	58369	58369

Note: Table shows regressions results from OLS on equation (2) in the main paper in the paper, where the dependent variable is the household expectation error in real income, winsorized at different thresholds (column 1: at 5% and 95% (the benchmark), column 2: 1% and 99%, column 3: 10% and 90%, column 4: 25% and 75%). The regressions included month dummies as additional controls. Standard errors reported in parentheses; they take the uncertainty induced by the imputation procedure into account.

B.2.5 Exclude observations with zero reported income change

Table 8: OLS of expectation errors on household characteristics, July only, observations with zero reported income change excluded

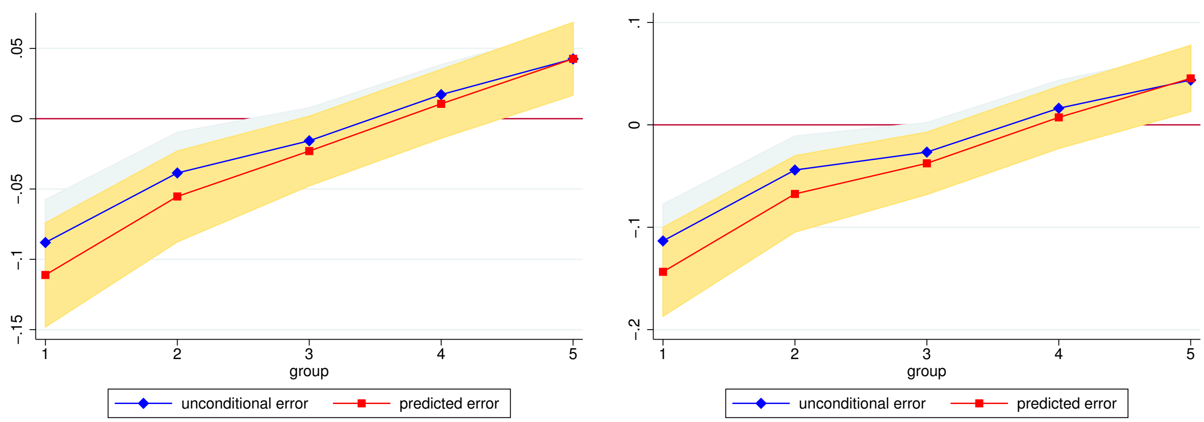
	(1) real	(2) real
Income Quintile		
1 (low)	−0.075	−0.091
	(0.021)	(0.025)
2	−0.038	−0.040
	(0.020)	(0.024)
4	0.025	0.032
	(0.016)	(0.020)
5 (high)	0.067	0.083
	(0.017)	(0.021)
Education		
educ=1	0.000	0.012
	(0.036)	(0.044)
educ=3	−0.032	−0.047
	(0.013)	(0.016)
Age		
age	−0.006	−0.008
	(0.004)	(0.005)
age × age	0.000	0.000
	(0.000)	(0.000)
Racial background		
black	0.021	0.023
	(0.022)	(0.025)
hispanic	0.018	0.022
	(0.033)	(0.038)
Number of adults		
1	0.026	0.031
	(0.042)	(0.047)
3 or more	0.021	0.028
	(0.022)	(0.027)
Other family characteristics		
female	−0.006	0.004
	(0.012)	(0.014)
not married	−0.019	−0.028
	(0.040)	(0.045)
Region		
North Central	−0.020	−0.021
	(0.017)	(0.021)
Northeast	−0.005	−0.006
	(0.018)	(0.022)
South	0.013	0.017
	(0.016)	(0.020)
Constant	0.132	0.163
	(0.094)	(0.110)
Sample	JULY	JULY
Observations	2805	2244

Note: The table shows results of the OLS regression of equation (2) in the main paper, where the dependent variable is the household expectation error in real income growth. Column 1 repeats the estimation on the full JULY sample (from table 1 in the main text), column 2 excludes observations which report no change in nominal income. The regressions include month dummies as additional controls. Heteroskedasticity-robust standard errors reported in parentheses.

Figure 4: Expectation errors in real income by income quintile, JULY sample, with and without observations that report zero income change

(a) all households

(b) exclude no change



Note: The figure shows the unconditional mean expectation error (blue line, diamonds) and predicted expectation error (red line, squares) by income quintile. Predicted expectation errors are based on regression results from table 8. Predicted values computed for all other explanatory variables at the weighted sample mean. Bands refer to 95% confidence intervals (based on heteroskedasticity-robust standard errors). On the y-axis, 0.01 corresponds to 1 percentage point.

B.3 Regression tables for error in aggregate unemployment expectation

Table 9: Ordered Logit / Ordered Probit of Unemployment Expectations

	(1) ologit	(2) oprobit
Income Quintile		
1st	−0.086	−0.046
	(0.023)	(0.013)
2nd	−0.032	−0.018
	(0.022)	(0.012)
4th	0.064	0.036
	(0.021)	(0.012)
5th	0.119	0.069
	(0.022)	(0.012)
Education		
no high school	−0.042	−0.017
	(0.039)	(0.022)
college	0.084	0.048
	(0.015)	(0.008)
Age		
age	−0.054	−0.031
	(0.005)	(0.003)
age × age	0.001	0.000
	(0.000)	(0.000)
Racial background		
black	−0.160	−0.074
	(0.029)	(0.016)
hispanic	0.078	0.051
	(0.035)	(0.020)
Number of adults		
1	−0.050	−0.025
	(0.030)	(0.017)
3 or more	0.083	0.048
	(0.024)	(0.014)
Other family characteristics		
female	−0.133	−0.084
	(0.014)	(0.008)
not married	−0.038	−0.024
	(0.028)	(0.016)
Region		
North Central	0.002	−0.002
	(0.020)	(0.011)
Northeast	−0.074	−0.041
	(0.022)	(0.012)
South	0.042	0.023
	(0.019)	(0.011)
Month dummies	yes	yes
Observations	96332	96332

Note: The table shows the results from the ordered logit and ordered probit regression of categorical errors in individual expectations about aggregate unemployment development. The ordered categories are as follows: -2: far too pessimistic, -1: too pessimistic, 0: correct expectation, +1: too optimistic, +2: far too optimistic. Heteroskedasticity-robust standard errors reported in parentheses.

B.4 Regression tables for actual and expected income growth

Table 10: OLS of growth expectations on observables

	(1) actual growth (real)	(2) actual growth (nominal)	(3) expected growth (real)	(4) expected growth (nominal)
Income Quintile				
1st	0.124	0.128	0.017	0.022
	(0.011)	(0.011)	(0.002)	(0.002)
2nd	0.052	0.054	0.006	0.009
	(0.009)	(0.010)	(0.002)	(0.002)
4th	−0.044	−0.045	−0.001	−0.003
	(0.007)	(0.008)	(0.002)	(0.002)
5th	−0.086	−0.089	0.003	−0.001
	(0.009)	(0.009)	(0.002)	(0.002)
Education				
no high school	−0.065	−0.067	−0.023	−0.019
	(0.017)	(0.017)	(0.003)	(0.003)
college	0.074	0.076	0.022	0.019
	(0.007)	(0.007)	(0.001)	(0.001)
Age				
age	0.007	0.007	−0.003	−0.003
	(0.002)	(0.002)	(0.000)	(0.000)
age × age	−0.000	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Racial background				
black	−0.052	−0.054	0.011	0.016
	(0.011)	(0.011)	(0.002)	(0.002)
hispanic	−0.034	−0.035	−0.005	0.002
	(0.013)	(0.013)	(0.003)	(0.003)
Number of adults				
1	0.077	0.078	0.001	0.003
	(0.020)	(0.021)	(0.003)	(0.003)
3 or more	−0.050	−0.052	0.003	0.001
	(0.010)	(0.011)	(0.002)	(0.002)
Other family characteristics				
female	−0.024	−0.025	−0.020	−0.013
	(0.006)	(0.006)	(0.001)	(0.001)
not married	−0.066	−0.067	0.010	0.009
	(0.018)	(0.019)	(0.003)	(0.003)
Region				
North Central	0.001	0.001	−0.018	−0.018
	(0.009)	(0.009)	(0.002)	(0.002)
Northeast	0.013	0.014	−0.011	−0.011
	(0.010)	(0.010)	(0.002)	(0.002)
South	0.005	0.006	−0.009	−0.008
	(0.009)	(0.009)	(0.002)	(0.002)
Constant	−0.082	−0.065	0.125	0.160
	(0.045)	(0.047)	(0.013)	(0.013)
Observations	18181	18181	89079	93764
R^2	0.039	0.040	0.046	0.047

Note: The table shows the results from the OLS regression of equation (2) in the main paper where the dependent variable is either actual income growth (columns 1 & 2) or expected growth (columns 3 & 4) in real or nominal income on the household level. Estimation for actual income growth performed on all households with re-interview; the regression includes year dummies as additional controls. Estimation for expected growth performed on full sample of households (with or without re-interview (first interview if there are two interviews), all interview months); the regression includes month dummies as additional controls. Heteroskedasticity-robust standard errors reported in parentheses.

C Alternative Mechanisms

In this section we go through alternative mechanisms that could potentially generate the same pattern of expectation errors. We argue that none of them is consistent with the empirical results.

Learning One potential explanation could be that people need to learn about their income potential over time, so that young households could be expected to make larger errors than older households. While in the regressions in the main text we already control for age effects, it might still be the case that expectation errors vary systematically with age. Figure 5 shows the unconditional as well as the predicted expectation errors for different age groups (holding all other characteristics, including income, at their sample mean). Panel (a) shows that the unconditional mean error is hump-shaped in age. However, once all other characteristics are controlled for, expectation errors are in fact decreasing with age, indicating that people become more and more pessimistic with age. It is not the case that expectations would improve as households age. Moreover, panel (b) shows that there is no clear pattern in inflation expectations with regards to age. Based on this result we conclude that people do not seem to learn about their income potential over time.

Inability to distinguish between persistent and transitory shocks In the income process typically considered in the literature there are two types of idiosyncratic shocks which differ in their persistence. The first type of shock is persistent. The other type is completely transitory. Could an inability to distinguish between the two shocks generate the pattern of expectation errors that we observe in the data? If households cannot tell the shocks apart and observe only overall income, they have to rely on some form of filtering to form beliefs about the current state. From linear projection theory we know that Kalman filtering is (conditionally) unbiased and optimal for linear systems and normal shocks. Hence there cannot be a systematic error conditional on past income developments if people form their beliefs optimally.

A sketch of a formal proof is the following. Consider a simple state space model

$$x_t = \rho x_{t-1} + \eta_t \tag{2}$$

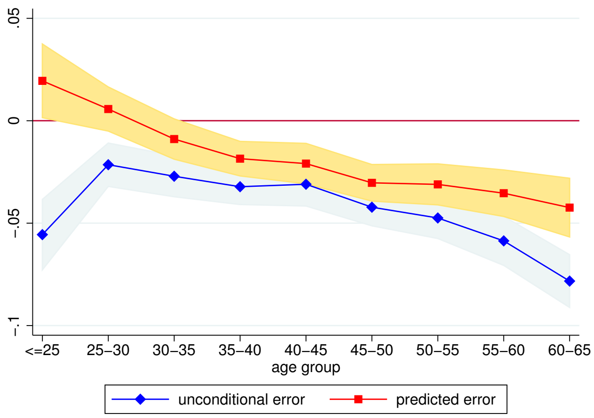
$$y_t = x_t + \mu_t \tag{3}$$

where η and μ are iid zero mean normal shocks with known finite variances. The forecasting error conditional on being in a particular quantile Q is $\mathsf{E}\left[y_{t+1} - y_{t+1|t} | y_t \in Q\right]$.

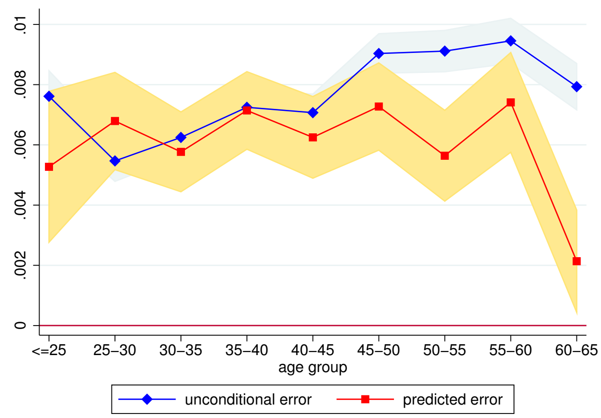
Suppose that t periods ago, the true state x_0 was known. It is then possible to write y_{t+1}

Figure 5: Expectation errors in real income by age group

(a) real income growth



(b) inflation



Note: The figure shows the unconditional mean expectation error (blue line, diamonds) and predicted expectation error (red line, squares) by income decile. Predicted expectation errors are based on regression results from table 1 in the main text, column 1 and 6, except that age is split into 5-year age groups instead of the quadratic term in age. Predicted values computed for all other explanatory variables at the weighted sample mean. Bands refer to 95% confidence intervals (for real income growth standard errors take the uncertainty induced by the imputation procedure into account; for inflation heteroskedasticity-robust standard errors are computed). On the y-axis, 0.01 corresponds to 1 percentage point.

as a function of starting state x_0, all previous η's and μ_{t+1}:

$$y_{t+1} = \eta_{t+1} + \mu_{t+1} + \rho\eta_t + \ldots + \rho^{t-1}\eta_1 + \rho^t x_0 \tag{4}$$

Similarly, $y_{t+1|t}$ can be written as a similar sum. However, now the noise terms μ also play a role because of imperfect information. It can be shown that

$$y_{t+1} - y_{t+1|t} = \eta_{t+1} + \mu_{t+1} + \rho\left[(1-K)\eta_t + K\mu_t\right] \tag{5}$$

where we assumed that the kalman gain K does not change over time.[1] The conditional forecasting error behaves similar to

$$\mathsf{E}\left[y_{t+1} - y_{t+1|t}|y_t\right] \approx \mathsf{E}\left[\eta_t - \mu_t \middle| \eta_t + \mu_t + \sum_{\tau=1}^{t-1} \rho^{t+1-\tau}\eta_\tau\right] \tag{6}$$

However, $\eta_t - \mu_t$ is independent of $\eta_t + \mu_t$ and because the shocks are not serially correlated, $\sum_{\tau=1}^{t-1} \rho^{t+1-\tau}\eta_\tau$ does not overturn the fact that the term in the expectations is independent of the condition. Hence the conditional forecasting error is equal to the unconditional, which is equal to zero.

[1]This approximation is better the bigger t is at exponential rate.

Table 11: Effect of Recent Experience on Growth Expectations

	(1) real	(2) real	(3) nominal	(4) nominal
past expectation	0.372	0.374	0.373	0.374
	(0.016)	(0.016)	(0.016)	(0.016)
past realized growth		−0.021		−0.022
		(0.004)		(0.004)
Income Quintile				
1st	0.004	0.007	0.007	0.009
	(0.004)	(0.004)	(0.004)	(0.004)
2nd	0.002	0.003	0.004	0.005
	(0.004)	(0.004)	(0.004)	(0.004)
4th	−0.005	−0.006	−0.005	−0.006
	(0.004)	(0.004)	(0.003)	(0.003)
5th	−0.008	−0.010	−0.008	−0.010
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.061	0.059	0.070	0.068
	(0.022)	(0.022)	(0.022)	(0.021)
Observations	15931	15931	17210	17210
R^2	0.185	0.187	0.182	0.184

Note: OLS estimation of individual growth expectations in 2nd interview as a function of past expectations and recent experience; estimation on sample 2HP (households with first interview in 2nd half of year and reinterview). Additional (unreported) control variables the same as in previous regressions: education, age, age^2, racial background, number of adults, gender, marriage status, region and time dummies. Heteroskedasticity-robust standard errors reported in parentheses.

Extrapolation from recent experience One explanation why current income can predict expectations about future income growth could be that people overweigh their recent experience. This would imply that households with a recent increase in income - which is correlated with being in a higher income group, all else equal - would expect another increase in the future.[2] We test for this explanation by regressing the growth expectations in the second interview on past expectations and recent experience (as well as on the other control variables we included in previous regressions). Table 11 shows that past expectations explain a large portion of current expectations, which means there is persistence in expectations on the individual level. The coefficient on recent experience, on the other hand, turns out to be significantly negative. This shows that households do not extrapolate from their recent experience. In fact, they seem to anticipate that there is mean reversion in their income process. Note, however, that the magnitude of this anticipated reversion is economically small. We can hence exclude extrapolation from recent experience as an explanation of the

[2] The relationship between expected income change and realized income change has been found to play a role in the analysis of Das and van Soest (1999).

systematic expectation errors by income groups.

Systematically wrong expectations about aggregates Another explanation for the observed pattern in expectation errors could be that households have biased expectations about aggregate conditions that vary systematically with their relative position in the income distribution. However, as seen in the analyses in the main text, household expectations about aggregate variables - such as inflation and the unemployment rate - are too pessimistic across the whole income distribution. Moreover, the magnitude of this bias doesn't vary much with income groups. Expectation errors in aggregate variables thus cannot explain the shift from overpessimism to overoptimism we observe as we move along the income distribution.

Measurement Error Since the empirical results are based on survey data we want to ensure that measurement error in reported variables is not the cause for the observed patterns. To do this we simulate an income process with persistent and transitory shocks as in the main text[3] and allow for four types of measurement error: errors in either the reported level income or the reported expectation in income growth, and each of these errors can either be an additive error or a multiplicative error. In detail, the information that is reported in the survey is assumed to have the following form:

$$\check{Y}_{it} = Y_{it} \cdot \xi^y_{it} + \varepsilon^y_{it} \tag{7}$$

$$\check{E}[g_{it}] = \frac{E[Y_{it+1}]}{Y_{it}} \cdot \xi^g_{it} + \varepsilon^g_{it} \tag{8}$$

$$\check{g}_{it} = \frac{\check{Y}_{it+1}}{\check{Y}_{it}} \tag{9}$$

where $\check{Y}_{it}$ and $\check{E}[g_{it}]$ are the reported income and reported growth expectations, respectively. $\check{g}_{it}$ is the realized income growth obtained from the reported level income. The additive measurement errors are normally distributed, the multiplicative errors log-normally:

$$\varepsilon^y_{it} \sim N(0, \sigma^y_\varepsilon) \tag{10}$$

$$\varepsilon^g_{it} \sim N(0, \sigma^g_\varepsilon) \tag{11}$$

$$\xi^y_{it} \sim \log N(-0.5(\sigma^y_\xi)^2, \sigma^y_\xi) \tag{12}$$

$$\xi^g_{it} \sim \log N(-0.5(\sigma^g_\xi)^2, \sigma^g_\xi) \tag{13}$$

[3]The income process is the same as employed in the main text. The difference is that we abstract from aggregate shocks and simulate the process directly on annual frequency.

We proceed by computing the observed forecast errors:

$$\check{\psi}_{it} = \check{E}[g]_{it} - \check{g}_{it} \tag{14}$$

We regress these errors by OLS on indicators for income quintiles, which are in turn determined based on reported income:

$$\check{\psi}_{it} = \alpha + \beta_1 \check{D}^1_{it} + \beta_2 \check{D}^2_{it} + \beta_4 \check{D}^4_{it} + \check{D}^5_{it} + \epsilon_{it} \tag{15}$$

Tables 12-15 show the resulting predicted forecast errors for increasing magnitudes of measurement errors in each of the four cases. The tables also show the distribution of measurement errors by income quintile and compare the magnitudes to the average income or growth rate in the respective income quintile.

Table 12 and table 13 show the results for measurement errors in reported level income. The considered magnitudes of these errors range from a standard deviation of 5% to 30% compared to the standard deviation of persistent income shocks. This translates into substantial measurement errors which are up to about 40% and 26% of mean income in the lowest income quintile for additive and multiplicative errors, respectively. Regarding the forecast errors that the OLS regression would predict, the tables show that the signs of these errors are broadly in line with the empirical findings. Quantitatively, however, even for large variances of measurement errors, the forecast errors are an order of magnitude smaller than what is found in the survey data. Table 14 and table 15 show that even for large measurement errors in reported expectations, there is no systematic effect on forecast errors.

We hence conclude that measurement errors in reported level income might contribute to the observed pattern, but they can at most explain a small fraction of the effects. Measurement errors in reported growth expectations do not contribute to predicted forecast errors.

Interpretation of survey answers as median or mode One potential concern about the way the Michigan Survey of Consumers phrases the expectation question is that it asks households for their expected income growth. But it does not explicitly specify whether households are supposed to supply their mean expected income growth, the median or even the mode. In the analysis in the main text we interpret their answers as the mean expected income growth.

However, even if households answered with the median or mode, this could not explain our systematic expectation errors. The reason is that the median and the mode of realized income growth rates are significantly lower than its mean for *both* low and high income households. Hence, if households answered with the median or mode of their expectations then both income groups would show up as too pessimistic in our analysis. This mechanism is therefore not able to explain why high income households are too optimistic – it is not able to generate a change in sign of the forecast errors.

Other mechanism Brunnermeier and Parker (2005) describe a setting where agents find it optimal to have too optimistic expectations. Alternatively, it might be possible that in order to attempt high risk-high reward projects, one needs to underestimate the chances of failure. The overoptimism for high income households could then arise as a result of survival bias. However, neither of these mechanism can explain why low income households are on average too pessimistic in their expectations. Regarding the low income agents, if there is ambiguity about the true income process, they might find it optimal to form expectations under a worst-case belief (Gilboa and Schmeidler, 1989; Epstein and Schneider, 2003). However, this mechanism cannot explain the overoptimism of high income households.

Table 12: Effects of Additive Measurement Errors in Reported Income Levels

$\sigma_{\varepsilon^y}/\sigma_P$	Predicted Forecast Errors			Distribution of Measurement Errors														
	Income Quintile			1st Income Quintile					3rd Income Quintile					5th Income Quintile				
	1st	3rd	5th	mean(Y)	min	5%	95%	max	mean(Y)	min	5%	95%	max	mean(Y)	min	5%	95%	max
0.05	-0.000	0.000	0.000	0.50	-0.03	-0.01	0.01	0.03	0.97	-0.04	-0.01	0.01	0.03	2.00	-0.03	-0.01	0.01	0.03
0.10	-0.002	0.000	0.000	0.50	-0.06	-0.02	0.02	0.06	0.97	-0.07	-0.02	0.02	0.06	2.00	-0.06	-0.02	0.02	0.07
0.15	-0.005	0.001	0.000	0.50	-0.10	-0.03	0.03	0.09	0.97	-0.11	-0.03	0.03	0.10	2.00	-0.09	-0.03	0.03	0.10
0.20	-0.009	0.001	0.001	0.50	-0.13	-0.04	0.04	0.12	0.97	-0.14	-0.04	0.04	0.13	2.00	-0.12	-0.04	0.04	0.13
0.25	-0.013	0.001	0.001	0.50	-0.17	-0.05	0.05	0.16	0.97	-0.18	-0.05	0.05	0.17	2.00	-0.15	-0.05	0.05	0.16
0.30	-0.019	0.002	0.001	0.50	-0.20	-0.06	0.05	0.18	0.97	-0.21	-0.06	0.06	0.20	2.00	-0.18	-0.06	0.06	0.20

Note: The table shows the effects of additive measurement errors in reported income levels. Column 1 shows the variation of the measurement error relative to the variation in persistent income shocks. Columns 2-4 show the the predicted forecast errors that result from running the main regression on simulated data. The remaining columns show the distribution of measurement errors for the different income quintiles and compare them to the mean income in the respective income quintile.

Table 13: Effects of Multiplicative Measurement Errors in Reported Income Levels

σ_{ξ^y}/σ_P	Predicted Forecast Errors Income Quintile			Distribution of Measurement Errors 1st Income Quintile					3rd Income Quintile					5th Income Quintile				
	1st	3rd	5th	mean(Y)	min	5%	95%	max	mean(Y)	min	5%	95%	max	mean(Y)	min	5%	95%	max
0.05	-0.000	0.000	0.000	0.50	-0.02	-0.01	0.00	0.02	0.97	-0.03	-0.01	0.01	0.04	2.00	-0.17	-0.02	0.02	0.19
0.10	-0.001	-0.000	0.000	0.50	-0.04	-0.01	0.01	0.04	0.97	-0.06	-0.02	0.02	0.06	2.00	-0.34	-0.04	0.04	0.38
0.15	-0.001	-0.000	0.001	0.50	-0.06	-0.02	0.01	0.05	0.97	-0.10	-0.03	0.03	0.09	2.00	-0.50	-0.06	0.06	0.58
0.20	-0.003	-0.000	0.001	0.50	-0.08	-0.02	0.02	0.07	0.97	-0.14	-0.04	0.04	0.12	2.00	-0.66	-0.08	0.08	0.78
0.25	-0.004	-0.001	0.002	0.50	-0.11	-0.03	0.02	0.09	0.97	-0.17	-0.05	0.05	0.15	2.00	-0.82	-0.10	0.10	0.97
0.30	-0.006	-0.001	0.002	0.50	-0.13	-0.03	0.03	0.10	0.97	-0.22	-0.06	0.06	0.18	2.00	-0.98	-0.11	0.12	1.18

Note: The table shows the effects of multiplicative measurement errors in reported income levels. Column 1 shows the variation of the measurement error relative to the variation in persistent income shocks. Columns 2-4 show the the predicted forecast errors that result from running the main regression on simulated data. The remaining columns show the distribution of measurement errors for the different income quintiles and compare them to the mean income in the respective income quintile.

Table 14: Effects of Additive Measurement Errors in Reported Growth Rate Expectations

$\frac{\sigma^2_{\varepsilon g}}{var(g)}$	Predicted Forecast Errors			Distribution of Measurement Errors														
	Income Quintile			1st Income Quintile					3rd Income Quintile					5th Income Quintile				
	1st	3rd	5th	mean(g)	min	5%	95%	max	mean(g)	min	5%	95%	max	mean(g)	min	5%	95%	max
0.01	0.000	0.000	0.000	1.30	-0.26	-0.07	0.07	0.22	1.06	-0.22	-0.07	0.07	0.22	0.87	-0.22	-0.07	0.07	0.23
0.02	0.000	0.000	0.000	1.30	-0.36	-0.10	0.10	0.31	1.06	-0.31	-0.10	0.10	0.31	0.87	-0.32	-0.10	0.10	0.33
0.03	0.000	0.000	0.000	1.30	-0.44	-0.12	0.12	0.37	1.06	-0.38	-0.12	0.12	0.38	0.87	-0.39	-0.12	0.12	0.40
0.04	0.000	0.000	0.000	1.30	-0.51	-0.14	0.14	0.43	1.06	-0.44	-0.14	0.14	0.43	0.87	-0.45	-0.14	0.14	0.46
0.05	0.000	0.000	0.000	1.30	-0.57	-0.15	0.15	0.48	1.06	-0.50	-0.15	0.15	0.49	0.87	-0.50	-0.15	0.15	0.52

Note: The table shows the effects of additive measurement errors in expected income growth. Column 1 shows the variation of the measurement error relative to the variation in income growth. Columns 2-4 show the the predicted forecast errors that result from running the main regression on simulated data. The remaining columns show the distribution of measurement errors for the different income quintiles and compare them to the mean growth rate in the respective income quintile.

Table 15: Effects of Multiplicative Measurement Errors in Reported Growth Rate Expectations

$\frac{var(\xi^g)}{var(g)}$	Predicted Forecast Errors			Distribution of Measurement Errors														
	Income Quintile			1st Income Quintile					3rd Income Quintile					5th Income Quintile				
	1st	3rd	5th	mean(g)	min	5%	95%	max	mean(g)	min	5%	95%	max	mean(g)	min	5%	95%	max
0.01	0.000	0.000	0.000	1.30	-0.43	-0.09	0.09	0.55	1.06	-0.33	-0.07	0.08	0.42	0.87	-0.35	-0.06	0.06	0.34
0.02	0.000	0.000	0.000	1.30	-0.59	-0.12	0.13	0.80	1.06	-0.44	-0.10	0.11	0.62	0.87	-0.48	-0.08	0.09	0.49
0.03	0.000	0.000	0.000	1.30	-0.71	-0.15	0.16	1.00	1.06	-0.53	-0.12	0.13	0.78	0.87	-0.58	-0.10	0.11	0.62
0.04	0.000	0.000	0.000	1.30	-0.81	-0.17	0.19	1.18	1.06	-0.60	-0.14	0.15	0.93	0.87	-0.65	-0.12	0.13	0.73
0.05	0.000	0.000	0.000	1.30	-0.90	-0.19	0.21	1.34	1.06	-0.65	-0.16	0.17	1.06	0.87	-0.72	-0.13	0.14	0.83

Note: The table shows the effects of multiplicative measurement errors in expected income growth. Column 1 shows the variation of the measurement error relative to the variation in income growth. Columns 2-4 show the the predicted forecast errors that result from running the main regression on simulated data. The remaining columns show the distribution of measurement errors for the different income quintiles and compare them to the mean growth rate in the respective income quintile.

D Mechanism: Overpersistence Bias - Proof of Theorem

D.1 Theorem

Income (net of age effects and the effects of other demographics) follows the process

$$Y_{it} = P_{it} \cdot T_{it} \tag{16}$$
$$P_{it} = P_{it-1}^{\rho} \cdot N_{it} \tag{17}$$

where P_{it} is a persistent component and T_{it} a transitory shock. Persistent income depends on past persistent income and a persistent shock N_{it}. Both shocks are independently and log-normally distributed with mean 1.

We assume that $1 > \hat{\rho} = \rho + \varepsilon > \rho$, so that all relevant moments exist and are finite. Expected income next period in this case is equal to $\mathsf{E}[Y_{it+1}] = \mathsf{E}[P_{it+1} \cdot T_{it+1}] = \mathsf{E}[P_{it}^{\hat{\rho}} \cdot N_{it+1} \cdot T_{it+1}] = P_{it}^{\hat{\rho}}$. Therefore the expected growth rate in income is $\mathsf{E}\left[\frac{\Delta Y_{it+1}}{Y_{it}}\right] = \frac{P_{it}^{\hat{\rho}} - Y_{it}}{Y_{it}}$ and the actual growth rate is equal to $\frac{\Delta Y_{it+1}}{Y_{it}} = \frac{P_{it}^{\rho} \cdot N_{it+1} \cdot T_{it+1} - Y_{it}}{Y_{it}}$ The expectation error can hence be calculated as:

$$\begin{aligned}
\psi_{it} &= E\left[\frac{\Delta Y_{it+1}}{Y_{it}}\right] - \frac{\Delta Y_{it+1}}{Y_{it}} \\
&= \frac{P_{it}^{\hat{\rho}} - Y_{it}}{Y_{it}} - \frac{P_{it}^{\rho} \cdot N_{it+1} \cdot T_{it+1} - Y_{it}}{Y_{it}} = \frac{P_{it}^{\hat{\rho}} - P_{it}^{\rho} \cdot N_{it+1} \cdot T_{it+1}}{Y_{it}} \\
&= \frac{P_{it}^{\rho+\varepsilon} - P_{it}^{\rho} \cdot N_{it+1} \cdot T_{it+1}}{Y_{it}} = \frac{P_{it}^{\rho}}{Y_{it}} \left(P_{it}^{\varepsilon} - N_{it+1} T_{it+1}\right) \\
&= \frac{P_{it}^{\rho-1}}{T_{it}} \left(P_{it}^{\varepsilon} - N_{it+1} T_{it+1}\right)
\end{aligned} \tag{18}$$

The *average* expectation error is then equal to $\mathsf{E}[\psi_{it}] = \frac{P_{it}^{\rho-1}}{T_{it}}\left[P_{it}^{\varepsilon} - 1\right]$. P_{it} can be re-written as a combination of its mean of $\mathsf{E}P = 1 + \bar{P}$ and the deviation from the mean p_{it}: $P_{it} = 1 + \bar{P} + p_{it}$. The term $\bar{P}$ is the lognormal mean correction term.

Using this notation, the expected error becomes

$$\mathsf{E}[\psi_{it}] = \frac{\left(1 + \bar{P} + p_{it}\right)^{\rho-1}}{T_{it}} \left[\left(1 + \bar{P} + p_{it}\right)^{\varepsilon} - 1\right] \tag{19}$$

For big enough current P_{it} (namely $p_{it} > -\bar{P}$), the term in the brackets is positive. This means that agents with income above this threshold on average overpredict their future income growth.

How does the expected error change with current P_{it}? $\frac{\partial \mathsf{E}\psi_{it}}{\partial P_{it}}$ has the same sign as $\frac{\partial F(z)}{\partial z}$ where $F(z) = z^{\rho+\varepsilon-1} - z^{\rho-1}$. We have

$$\begin{aligned} F(z)' &= z^{\rho-1}[(\rho+\varepsilon-1)z^{\varepsilon} - (\rho-1)] \\ &\approx (\rho+\varepsilon-1)\left[z^{\varepsilon} - \frac{\rho-1}{\rho-1+\varepsilon}\right] \\ &= -(1-\rho-\varepsilon)\left[z^{\varepsilon} - \frac{1-\rho}{1-\rho-\varepsilon}\right] \end{aligned} \tag{20}$$

This expression is *positive* as long as $z^{\varepsilon} < \frac{1-\rho}{1-\rho-\varepsilon}$, that is as long as $z < \left(\frac{1-\rho}{1-\rho-\varepsilon}\right)^{1/\varepsilon}$. Because $\rho \gg \varepsilon$ and ε is close to zero, the expectation error is increasing in P_{it} until very very large values of current P_{it}. In the model calibration, we have $\rho = 0.9774$, $\varepsilon = 0.0057$, which translates into a threshold of $z \approx 1.4e22$.

D.2 Corollary

If the true income process is governed by equations (3) and (4) in the main paper and the household overestimates the persistence of the process according to equation (5), the distorted expectation of next period's income is

$$\begin{aligned} E_t^{\theta}[\ln Y_{i,t+1}] &= \hat{\rho} \ln P_{i,t} \\ &= (\rho+\theta)\ln P_{i,t} \\ &= E_t[\ln Y_{i,t+1}] + \theta \cdot \sum_{s=0}^{\infty} \rho^s \ln N_{i,t-s} \\ &= E_t[\ln Y_{i,t+1}] + \theta \cdot \sum_{s=0}^{\infty} \rho^s \big(\ln P_{i,t-s} - E_{t-s-1}[\ln P_{i,t-s}]\big) \\ &= E_t[\ln Y_{i,t+1}] + \theta \cdot \sum_{s=0}^{\infty} \rho^{s-1} \big(\rho \ln P_{i,t-s} - \rho E_{t-s-1}[\ln P_{i,t-s}]\big) \\ &= E_t[\ln P_{i,t+1}] + \theta \cdot \sum_{s=0}^{\infty} \rho^{s-1} \big(E_{t-s}[\ln P_{i,t-s+1}] - E_{t-s-1}[\ln P_{i,t-s+1}]\big) \\ &= E_t[\ln Y_{i,t+1}] + \theta \cdot \sum_{s=0}^{\infty} \rho^{s-1} \big(E_{t-s}[\ln Y_{i,t-s+1}] - E_{t-s-1}[\ln Y_{i,t-s+1}]\big) \end{aligned}$$

Table 16: Mean expectation errors

	Data	Model			
ρ		0.9828	0.9880	0.9940	0.9990
$\hat{\rho}$		0.9868	0.9904	0.9949	0.9991
μ		0.9778	0.9778	0.9777	0.9776
income quintile 1	-0.0720	-0.0680	-0.0679	-0.0680	-0.0666
income quintile 2	-0.0370	-0.0409	-0.0409	-0.0385	-0.0396
income quintile 3	-0.0190	-0.0210	-0.0211	-0.0213	-0.0228
income quintile 4	-0.0000	-0.0013	-0.0014	-0.0057	-0.0050
income quintile 5	0.0160	0.0192	0.0193	0.0216	0.0220

*Note:*The results in this table show that the overpersistence bias can match the mean forecasting errors by income quintiles for varying values of the underlying persistence parameter.

E Overpersistence Bias for Alternative Income Process Parametrization

In the main text we use an income process parametrized according to Storesletten et al. (2004) as true underlying process. In this section we show that the expectation formation under overpersistence bias is able to match the empirically observed forecasting errors also for alternative specifications of the true underlying process. In particular, we show that the results are robust to increasing the persistence of the process and show how even the limit of random walk can be accommodated in our framework.

E.1 AR(1)

In this section, we keep the income process as a persistent AR(1) process with transitory shocks. We vary the persistence of the process and show that the overpersistence bias is able to match the patterns in the forecast error data very well even for very persistent processes.

This observation is documented in table 16. The first line shows the assumed persistence parameter of the underlying process (increasing from column to column). The second and third line show the two bias parameters that are needed to match the forecast errors. The last 5 lines show that for all assumed persistence parameters, the overpersistence bias is able to match the empirically observed forecast errors very well.

E.2 Random walk

In the limit case, the AR(1) becomes a random walk. In this section we show that a similar mechanism functions even in this setting. Here we assume that people do not observe the persistent and transitory component separately and have wrong beliefs about the relative volatility of the persistent component. This mechanism preserves the underlying intuition that people believe that shocks are more persistent than they truly are.

The agents receive income y_{it}, which is the combination of aggregate Z_t and idiosyncratic permanent P_{it} and transitory ε_{it}.

$$y_{it} = Z_t P_{it} \varepsilon_{it}, \tag{21}$$

$$P_{it} = P_{it-1} \eta_{it}, \tag{22}$$

where

$$\begin{bmatrix} \varepsilon \\ \eta \end{bmatrix} \sim \log \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \right) \tag{23}$$

People only observe y_{it}, not P_{it} and ε_{it} separately, so they use a Kalman filter to make forecasts about their future income. Following the same argument as in the main text, the households are allowed to make systematic mistakes when forecasting the aggregate component: $Z_{t+1|t} = \mu \mathsf{E}[Z_{t+1}]$. Because the aggregate bias does not change the dispersion of forecasting errors in the cross-section, we assume that $Z_t = 1$ for simplicity. Combined, the next period forecast is (dropping the individual index i):

$$y_{t+1|t} = \mu P_{t|t}, \tag{24}$$

where $P_{t|t}$ is the inferred permanent state based on the information set at time t. The resulting forecasting error in income growth is

$$\begin{aligned} error_t &= \log\left(\frac{\mu P_{t|t}}{P_t \varepsilon_t}\right) - \log\left(\frac{P_t \eta_{t+1} \varepsilon_{t+1}}{P_t \varepsilon_t}\right) \\ &= \log(\mu) + \log(P_{t|t}) - \log(P_t) - \log(\eta_{t+1}) - \log(\varepsilon_{t+1}) \end{aligned}$$

$\log(\eta_{t+1})$ and $\log(\varepsilon_{t+1})$ are on average zero, so the bias has to come from μ and the discrepancy between the believed and the actual state of the permanent income component.

The bias is implemented in the following way. The agents have wrong beliefs about σ_ε^2 and σ_η^2. This affects the Kalman gain, and changes how much of recent surprises in the observed y_t are believed to be caused by P as opposed to ε. We also assume that people are not wrong about the total variance $\sigma_\varepsilon^2 + \sigma_\eta^2$, so that people are correct about the one step

Table 17: Mean expectation errors

	Data	Model		
		Carroll et al. (2017)	Debacker et al. (2013)	Guvenen et al. (2016)
σ_ε		0.1000	0.3520	0.4940
σ_η		0.1000	0.0849	0.1560
γ_1		1.9994	1.5792	1.3661
μ		0.9779	0.9781	0.9783
income quintile 1	-0.0720	-0.0661	-0.0667	-0.0663
income quintile 2	-0.0370	-0.0427	-0.0405	-0.0417
income quintile 3	-0.0190	-0.0224	-0.0224	-0.0222
income quintile 4	-0.0000	-0.0017	-0.0039	-0.0026
income quintile 5	0.0160	0.0207	0.0215	0.0207

Note: The results in this table show that the overpersistence bias can match the mean forecasting errors by income quintiles even for the limit case of a random walk (for 3 different calibrations of the income process from the literature).

ahead conditional volatility of y. In other words, people believe that

$$\hat{\sigma}_\eta = \gamma_1 \sigma_\eta$$
$$\hat{\sigma}_\varepsilon = \gamma_2 \sigma_\varepsilon$$

such that

$$\gamma_1^2 \sigma_\eta^2 + \gamma_2^2 \sigma_\varepsilon^2 = \sigma_\eta^2 + \sigma_\varepsilon^2.$$

We simulate the income and forecast errors of 50,000 households over 50 years for 3 different parametrizations of the variances employed in the literature.[4] The results are captured in table 17. We show that this mechanism is able to match the empirical forecast errors for all 3 parametrizations of the random walk income process.

Note that qualitatively the mechanism always generates the correct direction of forecasting errors. In order to match the errors also quantitatively, it requires that there is enough volatility in the transitory shocks relative to the permanent shocks. The reason is that the bias falsely attributes parts of the realized transitory shocks to permanent shocks. For this to be able to generate large forecasting errors, sufficiently large transitory shocks are required to give enough scope for misallocation.

[4] Carroll et al. (2017) and Debacker et al. (2013) estimate the process on total household income which is the object of interest in our paper. Guvenen et al. (2016), on the other hand, only report estimates for total male labor earnings.

F Consumption Model: Numerical implementation

F.1 Solution Algorithm

The model is solved using a value function iteration algorithm with Howard's Improvement. The solution of the rational agent's problem is standard. The policy functions of the agent with biased bieliefs are obtained in two steps. First, the problem is solved using the grid and transition matrices as if the biased beliefs were correct. After the solution converges, we do one more iteration of the value function iteration algorithm, now using the grid corresponding to the true data generating process, keeping the transition matrices and the continuation values $\mathsf{E}V'$ from the biased agent solution.

Including the discretization of the aggregate and idiosyncratic income components, we solve the baseline model using the following grids:

- 210 grid points for liquid assets, unevenly spaced (step size gets smaller around zero) between around negative 11 and positive 12
- 120 grid points for durable assets, unevenly spaced (step size increasing with the level of durable asset) between 0 and 12
- 15 states for the persistent idiosyncratic component P, levels and transition matrices generated using Rouwenhorst method
- 11 states for the idiosyncratic transitory component T, levels and probabilities generated using Gauss-Hermite Quadrature
- 2 states for the aggregate component Z, calibrated so the model delivers the same time spent in booms and recessions as the US economy.

Presence of the durable adjustment costs implies that the household has to decide whether to incur these costs and choose the optimal level of durable asset or let the durable good depreciate. In theory, in each step of the value function iteration, the values for both action and inaction have to be updated. Solving for the optimal action given adjustment is particularly costly, because it involves two-dimensional optimization. However, in practice it is not necessary to update both value functions at all grid points. If one keeps track of the boundary of the inaction region, both values only need to be updated in the neighborhood of the boundary. This step can lower the solution time considerably for well chosen grids, as the inaction region will occupy a large fraction of the state space.

F.2 Simulation

We obtain the distributions by simulating a panel of 150000 households for 1500 periods (discarding the first 200 periods). Using the remaining 1300 periods, which include both booms and recessions as captured by the income component Z, we pool all the agents over all periods to construct the ergodic distributions.

To construct MPCs we run the following experiment. From the simulated series after discarding the burn-in period, we randomly select 300 points in time (so we correctly account for the effect of aggregate conditions) and use them as new starting points for new simulations. At these 300 starting points, we give every household a lump sum cash transfer and then we re-simulate the economy for one period. The MPCs are computed by comparing the behavior of households in the simulation with the transfer to the same households from the original simulation, averaged over all 150000 households and 300 simulations. This experiment uses all the memory that is available to us (256GB). In order to compute the impulse response functions, we are forced to limit the number of households to 30000 and simulate for 12 periods.

G Data Available in other Surveys

In this section we present the information about income expectations and realizations available in other household surveys and describe the challenges they pose for the analysis of rationality or biases of household income expectations. Compared to these other surveys, the Michigan Survey of Consumers in our opinion provides the best available data and allows the analysis of income expectations over a very long period of time.

G.1 Italian Survey of Income and Wealth (SHIW), Bank of Italy

The SHIW is a biannual panel household survey conducted by the Bank of Italy that has been running since 1977. Unfortunately, there is no overlap at all between income expectations and realizations.

Every other year the survey interviews the members of the participating households. It asks the following question about income expectations (see Bank of Italy (2018)):

> *B41 (ASPRED) This year, in 2017, do you expect your household's total income to rise more than prices, less than prices, or about the same as prices?*

For realizations, the survey asks in detail about different components of income of each household member (see sections B1-B6 in the questionnaire). However, all these questions refer to income in the calender year *prior* to the interview, while the expectation question refers to a change over the *next* calendar. Due to the biannual interview frequency that implies that there is no overlap at all between the expectations and the realizations (in addition to expectations being only qualitative). It is therefore not possible to use the SHIW to investigate rationality or biases in income expectations.

G.2 Survey of Consumer Expectations (SCE), Federal Reserve Bank of New York

The SCE is a survey that interviews a representative, rolling panel of US households about their economic expectations, including their income expectations. The survey started in 2013 and each household stays in the panel for at most 12 months. Unfortunately, income realizations are only elicited in bins which precludes the computation of income growth realizations.

In the core survey, the respondent is asked the following questions (see Federal Reserve Bank of New York (2018b)):

> *Q25v2 Next we would like to ask you about your overall household income going forward. By household we mean everyone who usually lives in your primary*

residence (including yourself), excluding roommates and renters. Over the next 12 months, what do you expect will happen to the total income of all members of your household (including you), from all sources before taxes and deductions? Over the next 12 months, I expect my total household income to...

- *increase by 0% or more*
- *decrease by 0% or more*

Q25v2part2 By about what percent do you expect your total household income to [increase/decrease as in Q25v2]? Please give your best guess.

Q47 Which category represents the total combined pre-tax income of all members of your household (including you) during the past 12 months?
Less than $10,000
$10,000 to $19,999
$20,000 to $29,999
$30,000 to $39,999
$40,000 to $49,999
$50,000 to $59,999
$60,000 to $74,999
$75,000 to $99,999
$100,000 to $149,999
$150,000 to $199,999
$200,000 or more

The problem is that the realization of household income is only reported in bins so that realized changes in income cannot be computed.

The survey also asks additional questions about the respondents labor earnings in their current job in their labor market survey (see Federal Reserve Bank of New York (2018a)). This survey is repeated every 4 months. The exact wording of the question is as follows:

L3 How much do you make before taxes and other deductions at your [main/current] job, on an annual basis? Please include any bonuses, overtime pay, tips or commissions.
... dollars per year

This question is answered in levels, not in bins. It asks the respondents to report the income they would have if they stayed a full year in the current job that they currently have. It does not ask about the earnings they actually had during the last year. If people change jobs or move in and out of unemployment these two objects can be very different. Question L3 thus refers to a different object than the expectation question Q25v2 (the expectation is about the income the household expects to earn in the next 12 months from any job they might work at during that period). Moreover, Q25 refers to total household income from all sources for all household members while L3 asks about labor earnings of the respondent only. Without further assumptions it is thus not possible to use question L3 as realization for expectations in question Q25. Since March 2015 the survey added another expectation question about annual earnings:

> *OO2e2 What do you believe your annual earnings will be in 4 months?*
> ... *dollars per year*

This question will be comparable to question L3 to analyze 4 month ahead expectations. Currently this question is only available for 2015/2016 so it is difficult to control for aggregate time effects. However, this source of data seems promising as more waves become available. D'Haultfoeuille et al. (2018) develop and exploit a new method of testing for rationality in this short sample and come to strikingly similar conclusions to the ones in this paper (even though the limited sample period does not allow them to control for aggregate effects).

G.3 Longitudinal Internet Studies for the Social Sciences (LISS), CentERdata

LISS is a panel survey that has been following a representative sample of Dutch households since October 2007. Unfortunately, only qualitative income expectations are solicited.

The survey asks the households to give categorical expectations about their "financial situation", not explicitly income expectations. In detail, the questions asked are the following (see CentERdata (2018)):

> *ci261 Do you expect your financial situation to get better or worse over the coming 12 months?*

- *will get much better*
- *will get slightly better*
- *will remain more or less the same*
- *will get a bit worse*
- *will get a lot worse*

ci243 Can you indicate, on a scale from 0 to 10, whether your financial situation has gotten better or worse compared to one year ago? 0 means that your financial situation has gotten much worse compared to one year ago 10 means that it has gotten much better.

These questions are certainly related to income expectations. However, their categorical nature and the fact that they mix income expectation into the general phrase of "financial situation" (which also subsumes other financial components such as expenditures, wealth developments, inheritances, etc) prohibits the direct analysis of income expectations.

G.4 Eurosystem's Household Finance and Consumption Survey (HFCS), European Central Bank

The HFCS collects data from household surveys in the Euro area. As of now there are two waves of this survey. Unfortunately, there is no overlap between income expectations and realizations

Most participating countries have repeated cross-section design which does not allow to connect realizations with expectations. Moreover, even for the few participating countries with a panel component (Belgium, Germany, Spain, Italy, Cyprus, Malta and the Netherlands, see Household Finance and Consumption Network (2017)) there is no overlap between the expectations an realization since the survey is conducted every 3 years but expectations are asked only for one year ahead. In detail, the question about income expectations the following (see Household Finance and Consumption Network (2012)):

7.13 HG0800 Over the next year, do you expect your (household's) total income to go up more than prices, less than prices, or about the same as prices?

- *More than prices*
- *Less than prices*
- *About the same as prices*

Income realizations are recorded as follows (as an example we show here the employee income component):

7.01A PG0100 Did you receive any sort of employee income during (last 12 months / last calendar year)?

7.01B PG0110 What was the total gross amount over (the last 12 months / last calendar year)? Please include income from regular wages or salaries, as *well as any overtime pay, tips, bonuses, profit sharing benefits (unless part of the pension arrangements).*

Thus, there is no overlap at all between the expectations ("over the next year") and the realizations since income realizations are asked 3 years later for "the last 12 months / last calendar year"[5]. Even if we were willing to work with the categorical expectations, the 3-year rhythm of the survey makes it therefore impossible to match realizations to the expectations.

References

Bank of Italy (2018). Household income and wealth - 2016 survey questionaire. downloaded on July 30, 2018 from https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/documentazione/index.html.

Barnard, J. and Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrica*, 86(4):948–955.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal expectations. *American Economic Review*, 95(4):1092–1118.

Carroll, C., Slacalek, J., Tokuoka, K., and White, M. N. (2017). The distribution of wealth and the marginal propensity to consume. *Quantitative Economics*, 8(3):977–1020.

CentERdata (2018). Liss panel data archive - 10 economic situation: Income. electronically accessed at https://www.dataarchive.lissdata.nl/study_units/view/35 on July 31, 2018.

Das, M. and van Soest, A. (1999). A panel data model for subjective information on household income growth. *Journal of Economic Behavior & Organization*, 40(4):409 – 426.

Debacker, J., Heim, B., Panousi, V., Rmnath, S., and Vidangos, I. (2013). Rising inequality: Transitory or persistent? new evidence from a panel of u.s. tax returns. *Brookings Papers on Economic Activity*, pages 67–142.

D'Haultfoeuille, X., Gaillac, C., and Maurel, A. (2018). Rationalizing rational expectations? tests and deviations.

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1 – 31.

Federal Reserve Bank of New York (2018a). Questionnaire for sce labor market survey. downloaded on July 30, 2018 from https://www.newyorkfed.org/medialibrary/media/research/microeconomics/interactive/downloads/sce-labor-questionnaire.pdf.

Federal Reserve Bank of New York (2018b). Survey of consumer expectations core questionaire. downloaded on July 30, 2018 from https://www.newyorkfed.org/medialibrary/interactives/sce/sce/downloads/data/FRBNY-SCE-Survey-Core-Module-Public-Questionnaire.pdf.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141 – 153.

[5]Note that the phrasing of last 12 months vs last calendar year differs between participating countries.

Guvenen, F., Kaharan, F., Ozkan, S., and Song, J. (2016). What do data on millions of u.s. workers reveal about life-cycle earnings dynamics?

Household Finance and Consumption Network (2012). Hfcs euro area questionaire wave 1. downloaded on July 31, 2018 from https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn.en.html.

Household Finance and Consumption Network (2017). The household finance and consumption survey: methodological report for the second wave. *ECB Statistics Paper Series no 17*.

Reiter, J. P. (2007). Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrica*, 94(2):502–508.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Cyclical Dynamics in Idiosyncratic Labor Market Risk. *Journal of Political Economy*, 112(3):695–717.