Subjective belief distributions in an Indian consumption panel with fisher households

Patrick Schmidt*
University of Zurich
Xavier Giné
World Bank

This version: February 10, 2022

Abstract

We investigate a panel of self-employed Indian fisher households with bimonthly income, subjective probabilities, and consumption data. We provide a rigorous framework to test rational expectations extending the Mincer-Zarnowitz regression to variance and probability integral transforms to analyze perceived uncertainty. A two-step instrumental variable approach can control for classical measurement error using multiple measurements of subjective probabilities. We find that subjective means are predictive, but reject full information rational expectations. Uncertainty statements are also informative, but to a much lesser degree and only for the male fishers, not for the female household heads. Adding to the literature on information frictions, we find evidence for information frictions within the fisher villages. Using the subjective probabilities, we try to understand the consumption response of the households. The evidence suggests that changes in expected income influence consumption, albeit less severe than a standard permanent income model would postulate. We disentangle permanent and temporary income shocks using expectations and find that permanent income changes are more important for consumption. Negative temporary income shocks also have some influence, which can indicate borrowing constraints. Finally, we confirm some results with a quasi-Bayesian GMM estimator that accounts for measurement error in expectations and consumption.

Keywords— subjective expectations, beliefs, consumption, income process
JEL-codes: E2, D8.

* *Address for correspondance:* E-mail: PatrickWolfgang.Schmidt@uzh.ch.

1 Introduction

The potential of subjective expectation data is increasingly leveraged for economic modeling (Manski, 2004). Point predictions have been used in a variety of studies to further our understanding of belief generation and economic decisions influenced by them. In this study, we try to advance this research agenda in several directions using panel data of Indian fisher households that contain income expectations and realizations as well as information on consumption.

In the first part of the paper, we focus on the subjective probabilities and analyze deviations from rational expectations under full information. In particular, we try to understand expectation formation and information frictions. The high-income uncertainty in developing countries in general, and for fishing income in particular, generates an economic environment where expectations may differ strongly across fishers and from the current income.

We show that the common pattern of information friction between households and overreaction in individual forecasts commonly observed for macro economic forecasts extends to the fisher households. Further, we extend Mincer-Zarnowitz(MZ) regressions from mean predictions to uncertainty statements (e.g., the variance). In contrast to the mean predictions, we find variance predictions to be less imprecise, while still in line with information frictions and overreaction.

It can be argued that measurement error is of particular importance for subjective probabilities, where direct validation is not feasible. For more complex aspects, like the variance of a subjective probability distribution, this might be even more important than for forecasts of central tendency (e.g., the mean or median). If the elicited probabilities are understood as a noisy measurement of a true underlying state of mind, research questions regarding belief generation and the impact of beliefs are best understood in terms of the unobserved underlying belief. To do so, we use a two-stage instrumental variable (IV) regression using a second, independent belief measurements as excluded instrument, which allows consistent inference under classical measurement error in Mincer-Zarnowitz regressions.

In the second part of the paper, we analyze consumption and saving decisions to uncover potential constraints that may arise in imperfect loan and insurance markets. We build on previous work using subjective probabilities from surveys in Italy (Pistaferri, 2001; Kaufmann and Pistaferri, 2009; Jappelli and Pistaferri, 2000), the United States (Attanasio et al., 2020), and Japan (Hayashi, 1985). The context of this study provides additional evidence on consumption smoothing and borrowing constraints in an environment with potentially different limitations in access to credits and informal insurance.

> *There is a widespread feeling that the wealthy have different motives to save from the less wealthy.* (Browning and Lusardi, 1996)

With this goal in mind, we focus on asymmetric responses to changes in income and expectations and find evidence for asymmetric effects indicating credit constraints.

The context of this study is of particular interest as the observed income process exhibits high volatility compared to wage labor or the survey populations in advanced economies. Additionally, deviations from rationality or consumption smoothing are especially costly for low-income households (Kueng, 2018). Thus, the welfare impact of consumption decisions may depend more heavily on dynamics between income, expectations, insurance, and borrowing constraints.

1.1 Methodological Contributions

In our analysis, we benefit from two features of the data concerning the subjective expectation data. As Arellano (2014) notes concerning subjective probabilities for income:

> *Progress [...] requires not only the input of researchers but also of data producing agencies.*

Two such contributions in data collection enable the study at hand: First, we have a non-rotating panel with 23 different measurement time points for many households, which allows us to consider household and/or time fixed effects in our analysis. Second, subjective expectations were elicited from the female and male household heads separately, which gives us two independent measurements of the beliefs of the household's decision makers.

The statistical approach taken here can be of interest beyond the application in this paper. In forecast evaluation, especially for survey forecasts with non-expert respondents, a two-stage IV approach with multiple measurements of subjective probabilities could be considered preferable as measurement error might otherwise invalidate inference. The generalizations of the MZ regression to the second moment and the probability integral transform (pit) might be of independent interest. The pit is defined as a function of a probability distribution and a realized outcome. It denotes the value of the cumulative distribution function at the realized outcome. The concept is well known in statistical forecast evaluation (Diebold et al., 1998; Gneiting and Ranjan, 2013) and might prove valuable in other economic contexts.[1] The pit is easily interpretable and the proposed first and second order pit rationality tests consider the whole distribution, which is often preferable to simply analyzing the mean.

We close the analysis of consumption responses with a moment-based approach based on Kaufmann and Pistaferri (2009). We alter the income process to accommodate a seasonal income process. We also propose a different estimator based on the Quasi-Bayes or Laplace type estimators proposed in Chernozhukov and Hong (2003) and Chen et al. (2018). Conveniently, the approach can handle weak or partial identification. The Markov chain Monte Carlo (MCMC) approach can be more stable than simple optimization in generalized method of moments (GMM) and maximum likelihood estimation, where local extreme points and non-smooth or non-convex target functions pose computational challenges.

1.2 Related Literature

Coibion and Gorodnichenko (2012) and Coibion and Gorodnichenko (2015) use a special case of our generalized MZ regression with forecast revisions as independent variable to investigate information rigidities and sensitivity to new information. The work inspired a quickly growing arm of research relying on related MZ regressions with similar findings for business expectations (Coibion et al., 2018; Bouchaud et al., 2019) and general macroeconomic forecasts (Bordalo et al., 2020; Kohlhas and Walther, 2021). For a more comprehensive review of applications see Bordalo et al. (2020).

[1] We note for example that Arellano et al. (2017) parameterize non-linear income dynamics in terms of the pit, which in their work is estimated by quantile regressions as no subjective distributions are available. See Arellano (2014) for an accessible introduction.

Using the first and second moment of the pit, we extend the MZ regression to the entire distribution. Other rationality tests for subjective probability distributions based on the pit have been proposed in the econometric literature (Rossi and Sekhposyan, 2019; Knüppel, 2015). In their work the focus lies on testing, whereas our regression-based approach aspires to describe the nature of deviation from full information rationality.

Attanasio (2009) discusses subjective probabilities from surveys in developing countries. Similar discussions can be found in Delavande et al. (2011b), who conclude that survey expectations in developing countries can be feasible and useful. Several studies use subjective expectations in the development context: Attanasio et al. (2019) elicit subjective distributions on revenues to study the impact of investment risks on loan demand. Thornton (2012) analyzes subjective beliefs about HIV infection status and economic responses to HIV test results in Malawi. Tarozzi et al. (2014) elicit the likelihood of catching Malaria with and without bed net use in India. The income dynamics in developing countries have been considered in Attanasio and Augsburg (2016) who find a persistent income process in rural India. Kaufmann (2014) and Attanasio and Kaufmann (2014) consider income expectations for school children to analyze education decisions in Mexico.

Income expectations have been considered in a series of papers (starting with Dominitz and Manski, 1997; Dominitz, 2001; Flavin, 1981). More recently, earnings expectations in the context of education choices were considered. Bunn et al. (2018) use survey data from the Bank of England that explicitly ask for unexpected income shocks and find negative shocks have a higher impact on consumption. Mostly this research focuses on subjective means to the exclusion of uncertainty measures like the variance (exceptions include Wiswall and Zafar, 2015; Wiswall and Zafar, 2021; Attanasio and Kaufmann, 2014).

The consumption response to income changes (marginal propensity to consume) has been analyzed based on survey evidence in many high-income countries. For heterogeneity results regarding household characteristics see for example Jappelli and Pistaferri (2020) and the references therein.

2 Data

The sample contains households from seven villages in the South Indian state of Tamil Nadu that are boat owners and have fishing as their main income source. Subsamples of these households were used to study labor supply decisions (Giné et al., 2017), recall bias in income surveys (De Nicola and Giné, 2014), elicitation designs for subjective expectations (Delavande et al., 2011a), and the validity of subjective expectations (Delavande et al., 2011b).

The data comes from four different sources. Daily book records of income from fishing, survey expectations for income elicited from male household heads, and a consumption and expectation survey with female household heads. Expectations and consumption were elicited every two months from 2011 until 2015. Finally, most households took part in at least one of two household surveys conducted in 2012 and 2013.

In each wave, the male and female household head were separately interviewed to elicit expectations. The survey of female household heads included a consumption survey after the expectation questions. Villages were visited in the same order in each survey wave. Surveys started two months before the target date for the expectation questions. The surveys in the last village took place about one month before the target date. As a consequence, the forecast horizon of expectation questions varied between 32 and 62 days.

The elicitation procedure was based on the best performing design in Delavande et al. (2011a). Respondents were asked to distribute 20 beans into a common income support divided into 20 intervals. Each bean represents a fixed probability such that the procedure identifies the cumulative distribution function of the subjective probability distribution at the interval thresholds (Eyting and Schmidt, 2021). For additional details on the elicitation see Delavande et al. (2011a).

Two different forecasting targets were applied. First, respondents were asked for "the catches in one day" and subsequently for the "monthly catches" in the target month. The target month was the month of the next survey round. Note that the two questions provide complementary information as the number of fishing days in a month may not be deterministic (Giné et al., 2017, shows that the days out fishing adapt to recent catches) and because daily catches exhibit autocorrelation (e.g., persistent weather, marine life condition, etc.).

Following Engelberg et al. (2009), we fit a parametric distribution to the elicited cumulative distribution function points ignoring rounding in the probabilities. From this parametric distribution, any property (e.g., mean, variance, standard deviation) can be computed. In particular, we fitted a normal distribution. Such approximation methods are necessary, as the survey tools do not identify the underlying distribution (compare e.g., Eyting and Schmidt, 2021). Note that seemingly non-parametric methods like computing the mean as $m = \sum_i p_i x_i$, where x_i denotes the midpoint of the interval and p_i the assigned probability, are essentially parametric fits with the assumption that distributions have zero mass beyond the midpoints of interval thresholds chosen by the researcher.

The consumption module answered by the female respondent elicited the value purchased in the last 30 days of over 300 specific categories. Throughout this study we consider aggregated non-durable consumption.

Households sell their daily catches to middlemen who record daily values in logbooks. We use these administrative data to construct daily and monthly catches. For additional details on the data generating procedure see De Nicola and Giné (2014) and references therein. The available data ranges from the year 2000 to 2015.

From the different data sources a panel data set was constructed, where one time step denotes two months. Such categorization is standard in the literature, but most consumption/income data is elicited in quarterly or yearly patterns. The most common assumption is that income and consumption elicited in a time period represent the information of the entire time period, and that expectations are informed by the entire time period. Both assumptions are most reasonable if the survey is administered at the end of the time period.

The situation at hand is more complicated as not all households were interviewed at the end of a month. Instead, the households were visited throughout the second month of each time period. As consumption was elicited for the past 30 days, the elicited value has a changing window of overlap with the assigned time point. Forecasts were elicited for a fixed target month irrespective of the timing of visit, which implies a different forecast horizon depending on the survey timing. Importantly, this also implies that the information set of households surveyed earlier is smaller as the current income for the respective month is not yet fully realized.

Income is available at a daily basis, which allows us to accommodate different income variables within our panel setting. In our data analysis we will also refer to income realized within 1, 7 or 30 days before the interview. Importantly, the income in the 30 days before the interview is not the same as the income in the month of the interview, but has some overlap depending on when in the month the interview was conducted. While the income before the interview can be in the information set of the forecast, the income after is not (unless it is fully predictable). See Christiano et al. (1991) for an example of the potential importance of time aggregation in analyzing the permanent income hypothesis.

After merging all data to an unbalanced panel data set, we have 3004 observations from 204 households with income, expectation, and consumption data. For additional details on the data see the supplementary material.

3 Statistical Evaluation of Subjective Probability Distributions

The relation between a subjective probability distribution and realized outcome is inherently difficult to measure. Each pair of distribution and outcome can be seen as a sample of a data generating process that includes both entities (see Gneiting and Katzfuss, 2014, for a review on probabilistic forecast evaluation). A single observation in such a sample can facilitate only limited inference without unduly strong assumptions on the underlying randomness of the process.

Early work considered point forecasts, as opposed to the probabilistic forecasts elicited in the study at hand, and tests if the forecast is a consistent mean forecast. The use of subjective distributions has the advantage that the predictive interpretation of the elicited content is not ambiguous. Simple point forecasts without explicit target functional reveal only ambiguous information (Schmidt et al., 2021; Engelberg et al., 2009). While untargeted point forecasts have been often understood as mean forecasts in economics, first evidence suggests that mode forecasts are more prominent in non-expert settings (Dimitriadis et al., 2019). In this study, we are able to generate the entire distribution from the survey instruments and extract any functional (e.g., the mean or the variance) from it. This avoids the ambiguity of unspecified point forecasts and allows us to extend rational expectation tests to the whole distribution.

We use the following notation. We denote the measured *income* for household i at time t by y_{it} often omitting one or all subscripts for brevity. Similarly, *consumption* is denoted by c_{it}. The first difference of a variable is denoted by $\Delta x_t = x_t - x_{t-1}$. The positive part of a variable x is denoted by $x^+ = \max(0, x)$.

Subjective probability distributions measured at time t with target income at time t' are denoted by $\mathbb{P}_{t,t'}$. We call the mean of the subjective probability distribution the *subjective mean (forecast)* $e_{t,t'}$. The subjective mean of the female household head is denoted by $e^f_{t',t}$. Often we omit the time of measurement for subjective probability variables. The same nomenclatures apply to the *subjective standard deviation* σ, *subjective variance* σ^2, etc. Note that we try to avoid referring to the subjective mean as "subjective expectation" as this is sometimes used for the entire subjective probability distribution (e.g. in Delavande et al., 2011b) or to the general concept of uncertainty perception (e.g. in Delavande, 2014).

We define *rationality* of the subjective distributions as the statistical consistency between the subjective probability distribution and the observation (Gneiting et al., 2007). A subjective distribution $\mathbb{P}$ is called rational for y, if $\mathbb{P}$ constitutes the conditional distribution of y given some information set. Statistical work often refers to this property (or a weaker notion thereof) as *calibration*. The conditional expectation of a random variable x with respect to the information available at time t is denoted by $\mathbb{E}_t[x]$.

In the following section, we begin by discussing the standard MZ type regressions for the subjective mean.

3.1 Standard Mincer-Zarnowitz Regression

A common procedure to test forecast rationality is a least squares regression of the realized outcome y on the subjective mean e:

$$y = \beta_0 + \beta_1 e + \epsilon.$$

A rational mean forecast would fulfill $E[y|e] = e$ which can be tested with

$$H_0 : \beta_0 = 0 \wedge \beta_1 = 1.$$

Such regressions are often called MZ regressions as they were pioneered in Mincer and Zarnowitz (1969) (for further discussions see Chong and Hendry, 1986). Note that by subtracting the subjective mean e, it can be seen that an equivalent model is the regression equation of the forecast error

$$y - e = \beta_0 + (\beta_1 - 1)e + \epsilon,$$

where rationality can be tested with $H_0 : \beta_0 = 0 \wedge \beta_1 - 1 = 0$. While this model is equivalent to the one presented above, it is noteworthy that the interpretation of the coefficient of determination R^2 differs (either giving the explained part of the outcome or of the forecast error). Further, a regression of the forecast error without including the subjective mean e as dependent variable (e.g., Coibion and Gorodnichenko, 2015; Bordalo et al., 2020) can suffer from omitted variable bias whenever the forecast predicts the forecast error. Also note that rational expectations imply that any variable in the information set $\mathcal{F}$ of the respondent is uncorrelated with the forecast error such that rationality implies the more general $\mathbb{E}[y - e|\mathcal{F}] = 0$, which can be tested with a general information sensitive regression

$$y = \beta_0 + \beta_1 e + \beta_x X + \epsilon$$

where the vector X includes potential information in the information set of the forecaster. A coefficient of $\beta_x \neq 0$ indicates that X is not fully accounted for in the forecast and rejects the joint hypothesis of rationality of the forecast e and of X being in the information set of the forecaster.

Throughout this section the income y_t is constructed by the sum of daily catches within the target month of the subjective mean $e_{t-1,t}$. Later sections will also consider the income y_t^c, which is constructed as the sum of daily catches in the 30 days that consumption was elicited for.

Table 1 shows results for the standard MZ regression. Here and throughout the paper we include the p-value for the Chi-squared test of the forecast coefficient being equal to one (denoted by "mean = 1") and of the joint hypothesis of rationality that also includes all other coefficients being equal to zero (denoted by "Rationality"). Column (1) and (2) contain the standard version with male and female subjective mean. In both cases we find subjective means to be highly predictive, but reject mean rationality. Column (3) and (4) establish that past income and last year income are important predictors with the latter explaining 52% of variation outperforming the subjective means. Columns (5) and (6) show the information sensitive MZ regression and suggest that male subjective means are informative beyond historic catches, but fail to include information generated by past catches.

An additional complication arises because of the error structure in the regression. Under the presence of common shocks, the difference between realized income and expectations are correlated. To control for common shocks between individuals at the same time we cluster standard errors at the time level. Table 2 repeats the previous analysis with standard errors clustered at the time

	y_t					
	(1)	(2)	(3)	(4)	(5)	(6)
$e_{t-1,t}$	0.487***				0.378***	0.100***
	(0.021)				(0.021)	(0.021)
$e^f_{t-1,t}$		0.498***				
		(0.021)				
y_{t-1}			0.403***		0.341***	
			(0.017)		(0.017)	
y_{t-6}				0.602***		0.568***
				(0.014)		(0.015)
Constant	20,881.060***	20,171.440***	28,038.890***	19,201.340***	11,239.790***	15,613.950***
	(1,087.971)	(1,126.972)	(804.301)	(648.022)	(1,194.889)	(992.576)
mean = 1 (pval)	$< 10^{-10}$	$< 10^{-10}$	-	-	$< 10^{-10}$	$< 10^{-10}$
Rationality (pval)	$< 10^{-10}$	$< 10^{-10}$	-	-	$< 10^{-10}$	$< 10^{-10}$
Observations	3,016	3,004	2,786	1,783	2,783	1,781
Adjusted R^2	0.153	0.152	0.165	0.519	0.254	0.527
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 1: Standard MZ regression. The subjective mean of the female household head is denoted by e^f.

level. We see that standard errors grow by the factor 2 to 3, indicating substantial correlation within waves. In this context, a cross-sectional analysis of a single wave (as often applied in the literature due to lack of a panel data set) can be heavy misleading without any means to control or estimate the impact of common shocks. In the remainder of this section MZ type regressions use covariance estimators that are clustered at the time level unless noted otherwise.

In this study, we consider two extensions to the classical MZ regression. First, we apply two-stage IV regressions to account for measurement error of expectations (for a review on measurement error see Hausman, 2001). Second, we extend the classical setting (for subjective means) to more general functions of the elicited distributions and the outcome to enable rationality analysis for uncertainty statements (e.g., for the subjective variance).

3.2 Two-stage Mincer-Zarnowitz Regressions

A potential reason for the rejection of rationality and the moderate explanatory power of the subjective means is measurement error in expectations. Consistent estimation in the presence of measurement error is possible with a two-stage IV approach first proposed in Jeong and Maddala (1991). For other approaches see Klepper and Leamer (1984) and Griliches and Hausman (1986). For many data sets with subjective probabilities, it is challenging to find strong instruments. In this study the availability of two different measurements of the subjective mean, from the male and female household head, allows an especially informative two-stage least squares regression. The two interviews were conducted on separate dates and in the absence of the other respondent. This makes the assumption of independent measurement error more credible.

In Table 3 we use the subjective mean of the other household head as excluded instrument. Compared to the standard least squares regression in Table 1 we obtain larger coefficients that are closer to 1, while still rejecting rationality. As before, the subjective means are found to be inconsistent with including all information from past income, while being informative beyond past income.

We note that Gillen et al. (2019) propose a similar approach for measurement error in experimental data, where multiple measurements are often available or can be easily included. They propose to combine both potential results (in our case the female households head instrumented by the male and vice versa) in a stacked two-stage regression to improve efficiency. In the case of experimental data, valid inference can be achieved by clustering standard errors at the individual level. As our observational data already has a complex error structure, we refrained from further complicating inference by stacking regressions and instead show the results side by side.

3.3 Heterogeneity in Predictive Performance

As the elicitation of subjective probabilities cannot be directly validated and is more challenging than most other economic data, their validity can be in doubt. One way to scrutinize the validity of an analysis based on subjective probabilities is to compare results between subsets that show signs of different accuracy. Here, we use the household specific ratio between the average squared forecast error of the subjective mean and the average variance of income. This measure can be interpreted as the mean squared error ratio between the subjective mean and the unconditional mean.

	y_t					
	(1)	(2)	(3)	(4)	(5)	(6)
$e_{t-1,t}$	0.487***				0.378***	0.100
	(0.089)				(0.097)	(0.066)
$e^f_{t-1,t}$		0.498***				
		(0.085)				
y_{t-1}			0.403***		0.341***	
			(0.054)		(0.049)	
y_{t-6}				0.602***		0.568***
				(0.045)		(0.046)
Constant	20,881.060***	20,171.440***	28,038.890***	19,201.340***	11,239.790**	15,613.950***
	(4,640.527)	(4,520.882)	(2,749.811)	(2,350.654)	(4,488.918)	(3,894.359)
mean = 1 (pval)	7e-09	3.6e-09	-	-	1.7e-10	$< 10^{-12}$
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	-	-	$< 10^{-12}$	$< 10^{-12}$
Observations	3,016	3,004	2,786	1,783	2,783	1,781
Adjusted R^2	0.153	0.152	0.165	0.519	0.254	0.527
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2: MZ regression with standard errors clustered at time level. The subjective mean of the female household head is denoted by e^f.

	y_t				
	(1)	(2)	(3)	(4)	(5)
y_{t-1}			0.316***		0.134**
			(0.049)		(0.062)
y_{t-6}				0.508***	0.474***
				(0.086)	(0.086)
$e_{t-1,t}$	0.663***		0.528***	0.273*	0.268*
	(0.113)		(0.140)	(0.145)	(0.148)
$e^f_{t-1,t}$		0.681***			
		(0.108)			
Constant	11,875.910**	10,690.700*	4,642.441	9,477.733*	5,139.337
	(6,015.459)	(5,701.664)	(6,400.071)	(5,049.173)	(4,593.461)
mean = 1 (pval)	0.0029	0.003	0.00078	5.8e-07	7.5e-07
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	3,004	3,004	2,774	1,777	1,774
Adjusted R^2	0.153	0.152	0.248	0.510	0.530
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 3: Two-stage IV MZ regression. As excluded instruments we apply the subjective mean of the female/male household head respectively. The subjective mean of the female household head is denoted by e^f.

Focusing on a subset of better performing boat owners (the best performing third of households) improves the performance as can be seen in Table 4. A mean coefficient of 1 is rejected only marginally after controlling via two-stage least squares in column (4) and (5). Full rationality, however, is still rejected decidedly.

We note that the joint hypothesis of rationality is strongly rejected, which cannot be deduced from the marginal distribution of effect estimates, but is instead driven by the estimated covariance matrix of coefficients. Further illustration of this mechanism can be found in Section 3.6.

Notably, the predictive performance of the models relying only on past income are not much better than in the full sample, which indicates that the more accurate subjective expectations are not the product of a more predictable income process.

	y_t					
	(1)	(2)	(3)	(4)	(5)	(6)
y_{t-6}					0.465^{***}	0.452^{***}
					(0.107)	(0.107)
y_{t-1}						0.121^{**}
						(0.059)
$e_{t-1,t}$	0.677^{***}		0.795^{***}		0.441^{**}	0.444^{**}
	(0.057)		(0.064)		(0.207)	(0.214)
$e^f_{t-1,t}$		0.615^{***}		0.869^{***}		
		(0.065)		(0.075)		
Constant	$12{,}972.660^{***}$	$16{,}089.650^{***}$	$6{,}786.034^{**}$	2,493.365	3,055.812	$-2{,}337.647$
	(2,806.046)	(3,295.207)	(3,436.184)	(3,987.077)	(9,041.683)	(8,385.691)
IV	No	No	Yes	Yes	Yes	Yes
mean = 1 (pval)	1.2e-08	4.1e-09	0.0013	0.081	0.0068	0.0095
Rationality (pval)	9.5e-11	9.4e-12	1.4e-09	1.5e-09	$< 10^{-12}$	$< 10^{-12}$
Observations	1,047	1,037	1,037	1,037	635	634
Adjusted R^2	0.278	0.230	0.278	0.230	0.530	0.541
Note:						$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: MZ regression with good performing households only. Results are based on the one third of households that explain the highest ratio of income variation by subjective mean of male household head. As excluded instruments we apply the subjective mean of the female/male household head respectively. The subjective mean of the female household head is denoted by e^f.

3.4 First Differenced Mincer-Zarnowitz regressions

The panel data setting allows additional rationality tests on innovations of the income process. Employing the fact that y_t is in the information set of the respondent at time t, it follows that

$$E_t[y_{t+1}] = e_{t,t+1} \iff E_t[\Delta y_{t+1}] = e_{t,t+1} - y_t,$$

which can be tested with a regression using the first difference of y_t as dependent variable.

Table 5 shows such regressions. In our setting only the income in the last 30 days before the survey y^c_{t-1} is certainly known to the household heads when issuing the forecast $e_{t-1,t}$. The income in the month of the survey y_{t-1} may not have fully realized yet (depending on the timing of the interview). Consequently, we construct the difference in the dependent variable with y^c_{t-1}.

Column (1) is directly derived from the equation above and rejects the hypothesis that the coefficient of y^c_{t-1} equals minus one. In the remaining columns, we use as dependent variable the difference $y_t - y^c_{t-1}$, accommodating the fact that the forecasting target is y_t, while the latest information available to the forecaster is y^c_{t-1}. While rational expectations are still rejected in column (2), we observe coefficients that are more in line with rationality and higher R^2 values indicating that subjective means are more suitable in predicting changes in income, than levels of income. If that holds true for expectation data in general, it can be argued that cross-sectional analysis is more vulnerable to inaccurate expectation data than panel analysis based on first differences.

In columns (3) to (5), we investigate which information is successfully integrated in the subjective probabilities and we try to find patterns of irrationality. Lagged income y^c_{t-1} (and lagged income innovation Δy^c_{t-1}) have a negative coefficient after controlling for the expected change $e_{t-1,t} - y^c_{t-1}$, indicating a overreaction to recent income changes. The income from last year y_{t-6} (and income innovation from last year $y_{t-6} - y^c_{t-7}$) have a positive coefficient, indicating a underreaction to this type of information. This observation is not consistent with rational expectations under full information, and instead suggests a recency bias (Erev and Haruvy, 2016) overweighting the recent past at the cost of the more distant.

3.5 Information in the Subjective Mean

It is interesting which information is rationally included in the subjective probabilities and what type of information is not consistently used in the expectation formation. In this section, we discuss this problem from a general perspective and look at fixed effects, forecast revisions, and consumption as potential information sources.

First we note that a non-zero coefficient of the forecast after including additional covariates implies that the forecast provides information beyond the used covariates (compare the forecast encompassing literature, e.g., Clements and Harvey, 2010). Second, a coefficient estimate of one for the subjective mean and zero for any other variable would suggest rationality with respect to all included covariates.

We further note that covariates that are irrelevant beyond the information rationally included in the forecast will have a population coefficient of zero irrespective of them being included in the forecasters information set. In other words, the test for the hypothesis of a rational forecast including a specific covariate as information has no power against the alternative that the specific

	Δy_t^c	y_t - y_{t-1}^c			
	(1)	(2)	(3)	(4)	(5)
$e_{t-1,t}$	0.507^{***}				
	(0.151)				
$y_{t-6} - y_{t-7}^c$				0.235^{***}	0.206^{**}
				(0.077)	(0.080)
y_{t-1}^c	-0.837^{***}	-0.211^{*}		-0.299^{**}	
	(0.068)	(0.110)		(0.126)	
y_{t-6}			0.280^{***}		
			(0.085)		
$\Delta\ y_{t-1}^c$					-0.139^{**}
					(0.061)
$e_{t-1,t}$ - y_{t-1}^c		0.608^{***}	0.806^{***}	0.446^{***}	0.603^{***}
		(0.127)	(0.071)	(0.162)	(0.105)
Constant	10,707.380	6,769.076	$-16{,}021.230^{***}$	$11{,}928.890^{*}$	$-2{,}407.432^{**}$
	(7,964.688)	(5,787.474)	(4,203.396)	(6,778.342)	(1,116.852)
mean = 1 (pval)	0.0011	0.0019	0.0064	0.00063	0.00015
Rationality (pval)	7.1e-11	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	2,764	2,970	1,776	1,744	1,741
R^2	0.426	0.499	0.594	0.636	0.615
Note:				*p<0.1; **p<0.05; ***p<0.01	

Table 5: Two-stage MZ regression with first differenced data. As excluded instruments we apply the subjective mean of the female/male household head respectively. The subjective mean of the female household head is denoted by e^f.

covariate is unknown to the forecaster but irrelevant beyond the information that is rationally employed by the forecaster.

Table 6 shows the two-stage MZ regression with different sets of covariates. The mean prediction remains statistically significant from zero after including past income and household fixed effects, indicating that the subjective mean contains information about income beyond those characteristics. The last years income y_{t-6} remains statistically significant different from zero after including fixed effects in column (4), suggesting insufficient adaptation to seasonal patterns. For this setting, the joint hypothesis of rationality cannot be rejected at the 5% level. We did not include the fixed effects in this hypothesis test. One can conclude that there is only weak evidence against rationality after accounting for individual specific biases. In column (6), after including the interaction of household fixed effects and monthly fixed effects for individual seasonal variation the estimated coefficient for the expectation is not statistically significant from zero anymore indicating no strong evidence for predictive performance beyond this aspect.

A particular interesting variable to construct information sensitive MZ regressions is the forecast, which can be easily argued to be in the information set of the forecaster. One such regression, based on the forecast revision, has obtained much attention recently (initiated by Coibion and Gorodnichenko, 2012; Coibion and Gorodnichenko, 2015). The literature often defines the forecast revision with respect to the same target date by $e_{t-1,t} - e_{t-2,t}$, which requires the existence of two-step ahead forecasts. Instead, we investigate MZ regressions based on the forecast revision of two forecasts with the same forecast horizon

$$\Delta e_{t-1,t} = e_{t-1,t} - e_{t-2,t-1},$$

which is unproblematic from a statistical point of view, as the crucial assumption that the revision is in the information set of the forecaster holds for both covariates. However, the interpretation of the regression results may differ. Under the assumption that $e_{t-2,t} = e_{t-2,t-1}$ both definitions are identical. Due to the seasonal pattern in fisher income this assumption is unrealistic for the process considered here.

In the following we try to shed light on the speed of adaptation to new information as well as on the complications that arise due to measurement error in this information sensitive MZ regression. Table 7 shows variants of the MZ regression with the forecast revision as independent variable. As before, rational forecasts would imply that the revision is not predictive after controlling for the actual forecast (or equivalently in the MZ regression with the forecast error as dependent variable that neither the forecast nor the revision are predictive covariates). It is common practice to regress the forecast error on the revision without controlling for the forecast itself (compare e.g., Coibion and Gorodnichenko, 2015; Bordalo et al., 2020) which can lead to an omitted variable bias, if the coefficient of the forecast deviates from rationality. This issue can arise if there exists measurement error in expectations, in which case the error term of a regression with dependent variable $y_t - e_{t-1,t}$ would be negatively correlated with $e_{t-1,t}$, leading to a downward bias of the coefficient estimate of $\Delta e_{t-1,t}$.

In column (1) we applied this potentially biased regression and observe a negative coefficient, indicating overreaction to new information, where an increase in expectations correlates with an lower than expected increase in the outcome. However, as we see in column (2) after controlling for the subjective mean there is no more indication of information frictions. Instead, there is evidence for the expectations $e_{t,t-1}$ coefficient to be different from zero.

	y_t					
	(1)	(2)	(3)	(4)	(5)	(6)
$e_{t-1,t}$	0.663***	0.528***	0.268*	0.439*	0.178	0.587
	(0.113)	(0.140)	(0.148)	(0.231)	(0.108)	(0.720)
y_{t-1}		0.316***	0.134**	0.005	0.153***	
		(0.049)	(0.062)	(0.072)	(0.043)	
y_{t-6}			0.474***	0.331**	0.495***	-0.099
			(0.086)	(0.133)	(0.064)	(0.216)
Constant	11,875.910**	4,642.441	5,139.337			18,034.210
	(6,015.459)	(6,400.071)	(4,593.461)			(27,585.710)
FE	-	-	-	hh	time	hh × month
mean = 1 (pval)	0.0029	0.00078	7.5e-07	0.015	$< 10^{-12}$	0.57
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	0.071	$< 10^{-12}$	8.2e-05
Observations	3,004	2,774	1,774	1,774	1,774	1,777
Adjusted R^2	0.153	0.248	0.530	0.522	0.629	0.585

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Two-stage MZ regression with covariates. As excluded instruments we apply the subjective mean of the female household head. The rationality test considers the restricted hypothesis of the shown coefficients being equal to zero excluding fixed effects in column (4) to (6).

	$y_t - e_{t-1,t}$				y_t			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$e_{t-1,t}$		-0.570^{***}	0.506^{***}	0.542^{***}	0.588^{***}			
		(0.086)	(0.069)	(0.109)	(0.110)			
$\Delta\ e_{t-1,t}$	-0.243^{***}	0.042	-0.251^{***}	-0.045	-0.217			
	(0.091)	(0.046)	(0.067)	(0.149)	(0.215)			
$\Delta\ e_{t-1,t}\ ^{+}$			0.463^{***}	0.312	0.364			
			(0.097)	(0.217)	(0.308)			
$e^f_{t-1,t}$						0.423^{***}	0.689^{***}	0.467^{***}
						(0.087)	(0.094)	(0.123)
$\Delta\ e^f_{t-1,t}$						-0.009	-0.290^{*}	0.428^{*}
						(0.081)	(0.169)	(0.243)
$\Delta\ e^f_{t-1,t}\ ^{+}$						0.153	0.616^{**}	-0.455
						(0.111)	(0.299)	(0.332)
IV	no	no	no	other hhh	daily	no	other hhh	daily
mean = 1 (pval)	-	$< 10^{-12}$	$< 10^{-12}$	2.5e-05	0.00017	4.3e-11	0.00093	1.5e-05
Rationality (pval)	1.4e-12	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	2,791	2,791	2,791	2,779	2,790	2,779	2,779	2,779
R^2	0.066	0.176	0.169	0.160	0.166	0.142	0.135	0.119

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: MZ regression with forecast revisions $\Delta e_{t-1,t}$. Constant terms are omitted. In columns marked in the IV row, we apply a two-stage IV procedure. As excluded instruments we apply the subjective mean of the female/male household head ("other hhh") or the subjective mean on daily income ("daily").

In column (3) to (5) we differentiate between positive and negative shocks. Column (3) reproduces the result in column (2), but shows strong evidence for an overreaction to negative and an underreaction to positive news. After using a two-step approach that can control for measurement error in expectations, the coefficients become smaller and statistically not significant at the 10%-level in column (4) with the excluded instrument being the other household heads expectations and of similar size in column (5) with the excluded instrument being the elicited daily expectations. We note that the latter cannot alleviate the bias from measurement error that influences the responses for daily and monthly expectations alike.

Column (6) to (8) mirror the three regressions before with the female household heads expectations. Results are similar, with the peculiarity that the two-stage IV regression in column (8), with the excluded instruments being constructed by the female daily income expectations, estimates no information rigidities, instead providing weak evidence for an overreaction to negative shocks. While the evidence is weak, the point merits discussion. One possibility is that negative news are only partly communicated by the fisher to the household (and the elicitor), and therefore the female household head tends, on average, to compensate by reacting overly sensitive to information that is not shared with or by the male fisher.

Table 8 shows a similar analysis with functions of the consumption as independent variable. As before, the consumption up to point t should be in the information set of the forecast (at least for the female expectations as the consumption was elicited from the same person at the same time). In column (1) and (4) we observe that consumption changes are negatively associated with the outcome after controlling for the subjective mean, irrespective of the household heads gender and the usage of instrumental variables for the subjective mean. Column (5) to (8) differentiate between consumption increases and decreases. Increases in consumption are associated with overly optimistic expectations, decreases show less association. However, throughout the evidence is weak as the error term exhibits a strong within-time correlation.

3.6 Mincer-Zarnowitz Variance Regressions

So far, we only considered the mean of the subjective probability distribution. A probability distribution provides additional aspects that can be evaluated, which may hold entirely different insights into the process of expectation formation. In particular, we would like to evaluate if the communicated uncertainty, often measured by the variance, standard deviation or an interquantile range, is rational given the data generating process. Systematic deviations from rationality, include overly narrow subjective distributions (in the following called overprecision, sometimes referred to as overconfident) or overly wide subjective distributions (in the following called underprecision). Extensions of the classical MZ regression to the second moment will allow us to analyze over- and underprecision marginally and their connection to covariates.

We consider two approaches to evaluate the forecasting performance of uncertainty perceptions. We begin by modeling the conditional variance $E[(Y - E[Y|\mathcal{F}])^2|\mathcal{F}]$. If the elicited mean e and variance σ^2 denote indeed the mean and variance with respect to some information set, it holds that $E[(Y-e)^2|\mathcal{F}] = \sigma^2$, which we test be regressing $(Y-e)^2$ on σ^2 and some additional variables that are potentially in the information set of the agent. As with the standard MZ regression, under rationality covariates in the information set $\mathcal{F}$ should not help explain the expected squared error and take a coefficient of zero in the population model.

	y_t							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$e_{t-1,t}$	0.468***	0.582***			0.464***	0.577***		
	(0.095)	(0.096)			(0.094)	(0.096)		
$e^f_{t-1,t}$			0.474***	0.621***			0.470***	0.625***
			(0.091)	(0.104)			(0.090)	(0.102)
$\Delta\ c_{t-1}\ ^+$					−0.626	−0.551	−0.694	−0.613
					(0.527)	(0.527)	(0.527)	(0.520)
$\Delta\ c_{t-1}$	−0.143	−0.166	−0.187	−0.231	0.197	0.133	0.189	0.099
	(0.235)	(0.225)	(0.246)	(0.234)	(0.427)	(0.420)	(0.424)	(0.411)
IV	no	daily	no	daily	no	daily	no	daily
mean = 1 (pval)	2.1e-08	1.2e-05	7.7e-09	0.00025	1.1e-08	9.9e-06	3.4e-09	0.00023
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	2,802	2,802	2,802	2,802	2,802	2,802	2,802	2,802
R^2	0.139	0.139	0.133	0.133	0.142	0.142	0.136	0.136
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 8: MZ regression with consumption revisions. In columns marked in the IV row, we apply a two-stage IV procedure. As excluded instruments we apply the subjective mean on daily income. The subjective mean of the female household head is denoted by e^f.

	$(\frac{y_t}{e_{t,t-1}}-1)^2$	$(\frac{y_t}{e^f_{t,t-1}}-1)^2$		$(\frac{y_t}{e_{t,t-1}}-1)^2$			$(\frac{y_t}{e^f_{t,t-1}}-1)^2$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta\ \frac{\sigma^2_{t-1,t}}{e^2_{t-1,t}}$					0.010	−0.038		
					(0.019)	(0.181)		
$\frac{\sigma^2_{t-1,t}}{e^2_{t-1,t}}$	0.289**		0.281**	0.053	0.333**	0.964*		
	(0.125)		(0.110)	(0.045)	(0.149)	(0.540)		
$\Delta\ \frac{\sigma^{2,f}_{t-1,t}}{e^{2,f}_{t-1,t}}$							0.029	0.047
							(0.037)	(0.131)
$\frac{\sigma^{2,f}_{t-1,t}}{e^{2,f}_{t-1,t}}$		0.104					0.102	0.843*
		(0.075)					(0.126)	(0.488)
Constant	0.035***	0.047***	0.055	0.010	0.029***	−0.025	0.044***	−0.014
	(0.009)	(0.006)	(0.040)	(0.008)	(0.009)	(0.044)	(0.009)	(0.038)
	-	-	hh FE	hh × month	-	IV	-	IV
variance = 1 (pval)	1.2e-08	$< 10^{-12}$	7e-11	$< 10^{-12}$	7.6e-06	0.95	$< 10^{-12}$	0.75
Rationality (pval)	2.5e-08	$< 10^{-12}$	2.8e-10	$< 10^{-12}$	5e-06	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	3,016	3,004	3,016	3,016	2,790	2,778	2,779	2,778
R^2	0.034	0.006	0.149	0.575	0.044	0.045	0.009	0.009
Adjusted R^2	0.034	0.006	0.086	0.283	0.044	0.044	0.008	0.008
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 9: MZ regression for variance. For regressions marked with “IV” we applied a two-stage IV approach, where the excluded instruments are the subjective mean of the female/male household head. The subjective standard deviation of the female household head is denoted by $\sigma^{2,f}$.

The MZ variance regression results can be seen in Table 9. Instead of the equation above, we standardize by e^2 and test an equivalent but computationally more convenient hypothesis

$$\tfrac{E[(Y-e)^2|\mathcal{F}]}{e^2} = E[(\tfrac{Y}{e} - 1)^2|\mathcal{F}] = \tfrac{\sigma^2}{e^2}.$$

In column (1) of Table 9 the subjective variance of the male household head is found to be a significant predictor, which suggests that the subjective variance contains information on the second moment of the underlying distribution. This evidence does not replicate for the female variance in column (2). Also note the difference in R^2 value, which denotes the explained variation in the second moment. Generally, a much lower ratio of the second moment of income is explained by the subjective probabilities than was the case for the first moment in a standard MZ regression.

In column (3), we control by household fixed effects. The positive coefficient of the male subjective variance suggests that the subjective statement contains information beyond the applied fixed effects. This pattern vanishes in column (4) after controlling for individual seasonal specific fixed effects. Thus, there is no evidence that the subjective variance is moved by additional information beyond seasonal adaptation.

Similar to the analysis of the subjective mean, the MZ variance regression using two-stage IV, that provides consistent estimates under measurement error, rejects a fully rational subjective variance. In contrast to the subjective mean analysis, the IV regressions in column (6) and (8) cannot reject a coefficient of 1 for the subjective variance. Note that rationality is often rejected decisively, while single coefficient estimates do not contradict rationality. This can be explained by the estimated covariance. If deviations from rationality are in the same direction, but the covariance estimate signals a negative correlation, the pattern arises. Consider as example the column (6) in Table 9. Here, none of the coefficients deviate from rationality if a Wald-test is applied to each coefficient separately. Yet, the joined rationality test is decidedly rejected with a p-value below 10^{-12}. Figure 1 illustrates that the negative correlation between the estimates for the variance and the constant are indeed consistent with such a discrepancy.

3.7 Mincer-Zarnowitz PIT-Regressions

While the rationality of point forecasts could be assessed with a simple analysis of the forecast error, subjective probability distributions require a more sophisticated approach. We propose to analyze rationality with the first and second moment of the Probability Integral Transform (pit), where the first moment reflects bias in position of the distribution (optimism and pessimism) and the second moment in spread of the distribution (overconfidence and underconfidence). The pit is defined as the probability that the subjective distribution $\mathbb{P}$ assigns to the interval below the realized outcome y:

$$pit_t = \mathbb{P}_{t-1,t}((-\infty, y_t]).$$

If the subjective probability distributions are rational, the pit is uniformly distributed between zero and one (Diebold et al., 1998; Gneiting and Ranjan, 2013). If the outcome is consistently larger than postulated by the subjective probability distribution, the average pit is close to one. If the outcome is consistently smaller, the average pit is closer to zero. We call a sample of subjective probability distributions *first order rational* if the average pit is equal to $\frac{1}{2}$. First order rationality is a necessary (but not sufficient) condition for rationality.

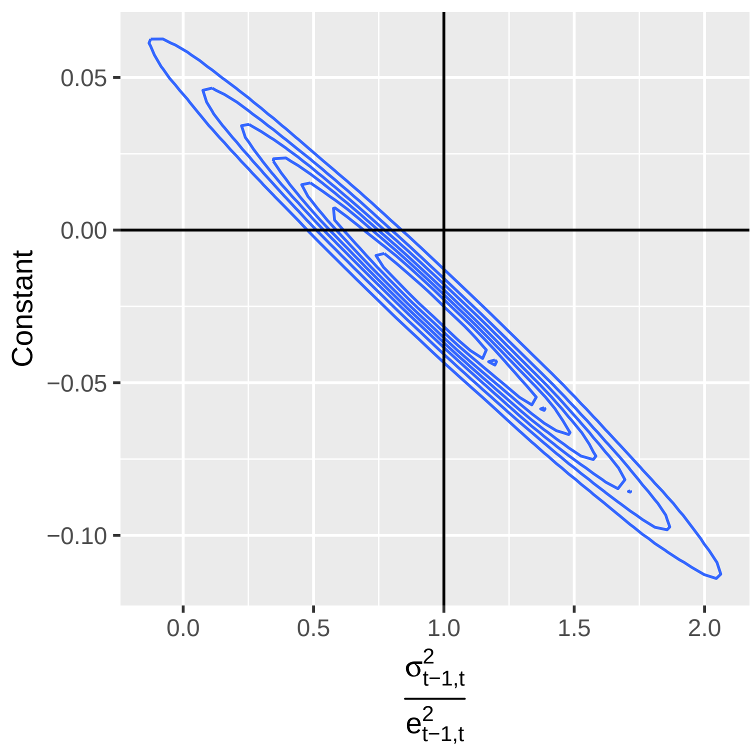

Figure 1: Density plot based on 1000 draws from the asymptotic multivariate normal distribution estimated in column (6) in Table 9.

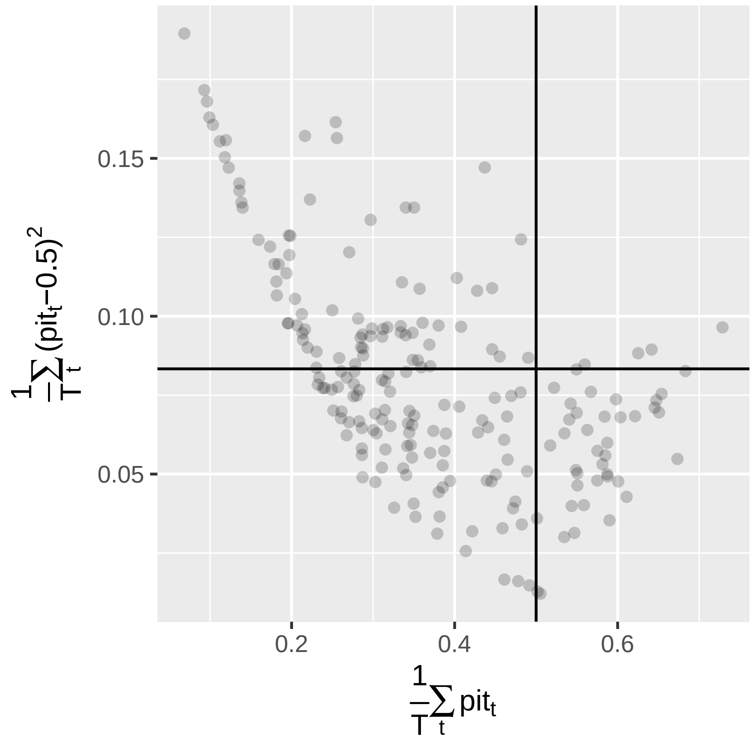

Figure 2: Empirical mean and variance of pit for all boat owners.

A uniformly distributed pit has variance $\frac{1}{12}$. We call a sample of subjective distributions *second order rational* if the variance of the pit is equal to $\frac{1}{12}$. Predictive distributions that are overconfident (too narrow) have a larger pit variance, subjective probability distributions with an overly wide spread have smaller pit variance.

In Figure 2 we see empirical moments of pit by household. The first moment is consistently below 0.5 indicating overly optimistic income expectations, which can complicate the analysis of the second moment. For those boat owners with realistic first order expectations, the second moment is moderately lower than 1/12 indicating underprecision (expectations indicate more variation than observed).

We note that the two functions of the pit, $pit - 0.5$ and $(pit - 0.5)^2 - \frac{1}{12}$, resemble "generalized forecast errors" in the sense of Patton and Timmermann (2007) but aimed at probability forecast instead of point forecasts. See also Schmidt et al. (2021) for interpretations of point forecasts, where the concept is referred to as identification function.

To test if deviations are statistically significant, we propose a regression analysis with transforms of the pit as dependent variable. First, bias is analyzed in Table 10 with $pit - 0.5$ as dependent variable. Under rational expectations, the conditional expectation of the dependent variable is zero and covariates known at the time of the forecast exhibit are uncorrelated. Negative values indicate realizations that are lower than the subjective distributions, positive values indicate higher than expected values. As income shocks can be expected to be correlated across time, the following regressions are clustered at the time level.

Deviations from rationality are clearly detectable. We observe a negative constant, indicating overly optimistic expectations. Additional evidence for limited information arises from the coefficient estimates. In column (1), the forecast revision and the lagged pit are found to be predictive. With additional controls in column (2), this effect vanishes. Instead, we see that the deviations from rationality are mostly driven by underreaction to the information from last year and of the subjective mean being overly sensitive, i.e. high values indicating a lower than expected income and vice versa. It is noteworthy, that this over-sensitivity would be attributed mostly to information rigidity Δe in regressions not controlling for the subjective mean e. In column (3), we add labor supply (the numbers of days out fishing) as independent variable. The main conclusions from above hold. The non-zero coefficient of labor suggests that fishers cannot fully predict their labor supply two months ahead. In particular, the positive coefficient indicates that additional fishing cannot compensate in months with weak income, instead driving some of the unexpected variation. Column (4) and (5) repeat the analysis with household fixed effects supporting the aforementioned conclusions. Columns (6) to (8) use individual specific seasonal fixed effects. Again, similar conclusions can be drawn. For the seasonally adjusted deviations of the pit, there is no evidence in column (8) that the last years income is not sufficiently adjusted for.

Noteworthy, rationality tests in column (4), (6), and (7) without labor supply and allowing for non-zero fixed effects cannot reject this limited form of rational expectations convincingly. If labor supply is included, rationality is rejected decidedly with respect to this larger information set, indicating that labor supply is not deterministic given the information available to the fisher.

A similar analysis of the variance of the pit can be used to analyze if subjective probability distributions over- or underpredict uncertainty. If the outcome is consistently more extreme (smaller or larger) than postulated by the subjective probability distribution, the pit is more likely to be close to either zero or one. This indicates overprecision of the subjective probability distribution. If the outcome is consistently less extreme, the pit is on average closer to 0.5. This indicates underprecision.

Table 11 shows results of the second moment pit MZ regression. In column (1) without additional controls, we find the squared deviation of the lagged pit $(pit_{t-1} - 0.5)^2$ to be predictive, but not the innovation in the subjective variance $\Delta\ \sigma^2_{t-1,t}$. After adding the subjective variance and past deviations from the seasonal average in column (2), the standard error estimates increase and the evidence becomes less conclusive. Instead, deviations from last year have a positive effect, indicating that subjective probabilities react insufficiently to this information. In column (3), we add the deviation of labor supply and find no influence in contrast to the previous findings on first order pit, where labor supply did explain some of the pit variation. The adjusted R^2 moves only marginally indicating that labor supply explains only a minor part of income uncertainty. Through the constant term, columns (1) to (3) also provide some evidence that the empirical variance of the pit is lower than that of a uniform distribution, implying that outcomes are closer to the center of the subjective probability distributions than under rationality, which suggests under-precision (overly wide subjective probability distributions). However, the evidence is weak and a bias of zero cannot be rejected at the 10%-level. Noteworthy, this tendency to underprecision stands in contrast to the finding that subjective probabilities tend to overprecision (i.e, overly confident statements with too narrow distributions) (Moore and Healy, 2008).

After adding household fixed effects in column (4) and (5) the findings persist but evidence gets

	$pit_t - 0.5$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta\ e_{t-1,t}\ (s)$	-0.058^{**}	-0.008	-0.016	-0.026	0.030	-0.067	0.005	-0.019
	(0.026)	(0.053)	(0.054)	(0.026)	(0.045)	(0.114)	(0.079)	(0.106)
$pit_{t-1}\ (s)$	0.117^{***}	0.057	0.052	0.022	-0.015	0.056^{*}	0.032	0.073
	(0.012)	(0.055)	(0.060)	(0.017)	(0.051)	(0.030)	(0.026)	(0.107)
$e_{t-1,t}\ (s)$		-0.133^{***}	-0.150^{***}		-0.158^{***}		-0.119^{*}	-0.158^{*}
		(0.048)	(0.045)		(0.043)		(0.070)	(0.090)
$labor_{t-1}\ (s)$			-0.071^{**}		-0.015			-0.014
			(0.032)		(0.031)			(0.025)
$labor_t\ (s)$			0.138^{***}		0.156^{***}			0.170^{***}
			(0.029)		(0.032)			(0.052)
$y_{t-1}\ (s)$		-0.002	0.014		0.019			-0.057
		(0.053)	(0.062)		(0.048)			(0.097)
$y_{t-6}\ (s)$		0.122^{***}	0.113^{***}		0.077^{***}			-0.026
		(0.028)	(0.024)		(0.028)			(0.028)
Constant	-0.127^{***}	-0.115^{***}	-0.133^{***}					
	(0.018)	(0.015)	(0.019)					
FE	-	-	-	hh	hh	hh × month	hh × month	hh × month
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	0.42	$< 10^{-12}$	0.12	0.051	$< 10^{-12}$
Observations	2,772	1,774	1,774	2,772	1,774	2,772	2,772	1,774
Adjusted R^2	0.171	0.458	0.526	0.303	0.553	0.499	0.576	0.679
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 10: Two-stage IV-regression for first moment of pit. As excluded instruments we apply the subjective mean of the female/male household head respectively and according transforms for first differences and the pit. *Labor* denotes the number of days out fishing. Standard errors are clustered at the time level. The rationality test considers the restricted hypothesis of the shown coefficients being equal to zero excluding fixed effects.

	$(pit_t - 0.5)^2 - \frac{1}{12}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta\ \sigma^2_{t-1,t}$ (s)	−0.068	−0.094	−0.096	−0.051	−0.091	0.013	0.007	−0.052
	(0.043)	(0.098)	(0.100)	(0.039)	(0.098)	(0.123)	(0.155)	(0.093)
$(pit_{t-1} - 0.5)^2$ (s)	0.022***	0.029	0.029	0.005	0.020	0.009	0.008	0.032
	(0.005)	(0.030)	(0.031)	(0.006)	(0.027)	(0.012)	(0.012)	(0.020)
$\sigma^2_{t-1,t}$ (s)		−0.001	−0.011		0.009		0.010	0.017
		(0.135)	(0.146)		(0.101)		(0.126)	(0.132)
$\lvert labor_{t-1} - \overline{labor}_{t-1}\rvert$ (s)			0.002		0.009			−0.005
			(0.006)		(0.006)			(0.010)
$\lvert labor_t - \overline{labor}_t\rvert$ (s)			0.007		0.012			0.017
			(0.010)		(0.012)			(0.016)
$\lvert y_{t-1} - \overline{y}_{t-1}\rvert$ (s)		−0.019	−0.020		−0.023			−0.016
		(0.028)	(0.028)		(0.028)			(0.014)
$\lvert y_{t-6} - \overline{y}_{t-6}\rvert$ (s)		0.018***	0.018***		0.011			−0.013***
		(0.007)	(0.006)		(0.007)			(0.005)
Constant	−0.005	−0.013	−0.012					
	(0.004)	(0.010)	(0.011)					
FE	-	-	-	hh	hh	hh × month	hh × month	hh × month
Rationality (pval)	9.3e-09	1.6e-06	$< 10^{-12}$	0.44	0.012	0.13	0.25	4.3e-06
Observations	2,772	1,774	1,774	2,772	1,774	2,772	2,772	1,774
Adjusted R^2	0.104	0.102	0.111	0.136	0.048	0.008	−0.041	0.161
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 11: Two-stage IV-regression for second moment of pit. As excluded instruments we apply the subjective mean of the female/male household head respectively and according transforms for first differences and the pit. *Labor* denotes the number of days out fishing. Here, $\overline{x}_t$ indicates the empirical average of x for the same month and household. *Labor* denotes the number of days out fishing. Standard errors are clustered at the time level. Standard errors are clustered at the time level. The rationality test considers the restricted hypothesis of the shown coefficients being equal to zero excluding fixed effects.

non-conclusive. After controlling with individual specific seasonal fixed effects in column (6) to (8), seasonal effects are accounted for. In this case, we observe an overreaction to the information in the deviation from last year, which suggests that the respondents under-react to the seasonal information from last years observation, but overreact to the non-seasonal part of it.

Rationality is clearly rejected in column (1) to (3), where the constant is tested. In column (4) to (8), where the fixed effects are excluded from the hypothesis and a less restrictive hypothesis is tested for, we observe rejections of rationality only in column (5) and (8) where additional information is included. The more standard specifications in column (4) and (6) do not reject rationality with respect to the included information.

3.8 Information Friction and Overreaction

In this section, we combine the introduced methods to investigate two widespread documented phenomena: overreaction to individual news and information frictions. The analysis is inspired by the work on macro economic forecasting (for further references see Bordalo et al., 2020). We bring the three technical innovations from the previous sections to bear: Two-stage IV with a second measurement, inclusion of the forecast as independent variable, and uncertainty focused MZ regressions.

In Table 12 and 13 we analyze if there exist overreactions or information frictions for central forecast (mean and first order pit) and for uncertainty (variance and second order pit) separately. In both tables the first two columns consider male expectations (instrumenting with female expectations) and columns (3) and (4) consider female expectations (instrumenting with male expectations). Column (1) and (3) are based on the mean and variance of the forecast error. Column (2) and (4) are based on the first and second moment of the pit.

In Table 12, we use the forecast $e_{t-1,t}$, the individual forecast revision $\Delta e_{t-1,t}$ and the average forecast revision within a village $\Delta e^v_{t-1,t}$ as independent variables. If the information shared during the survey is shared within a community without frictions, the village revision should not be predictive. Instead, we observe a positive coefficient for the village revision indicating that information within the village is not fully incorporated and individual expectations do not sufficiently adapt to new information available among members of the village. Further, we estimate a negative coefficient for the individual revision, indicating overreaction to new information. The latter effect is not statistically significant from zero for the male household head.

Similarly, in Table 13, we use the forecast

$$x_t = \frac{\sigma^2_{t-1,t}}{e^2_{t-1,t}},$$

the individual forecast revision Δx_t and the average forecast revision within a village Δx^v_t as independent variables. As before, we observe a positive coefficient for the village revision indicating the same information friction for the uncertainty faced by the households. The individual revision is also negative, indicating overreaction to new information. However, the latter effect is not statistically significant from zero for both household heads.

	y_t - $e_{t-1,t}$	$pit_t - 0.5$	y_t - $e^f_{t-1,t}$	$pit^f_t - 0.5$
	(1)	(2)	(3)	(4)
forecast	-0.476^{***}	-0.093^{***}	-0.387^{***}	-0.070^{***}
	(0.131)	(0.024)	(0.131)	(0.026)
village revision	0.362^{**}	0.077^{**}	0.422^{**}	0.080^{**}
	(0.157)	(0.037)	(0.187)	(0.032)
individual revision	-0.167	-0.040	-0.272^{*}	-0.063^{*}
	(0.146)	(0.041)	(0.154)	(0.034)
Constant	$19{,}202.520^{***}$	-0.123^{***}	$14{,}368.230^{**}$	-0.133^{***}
	(6,978.386)	(0.017)	(6,986.468)	(0.016)
Rationality (pval)	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$	$< 10^{-12}$
Observations	2,779	2,779	2,779	2,779
Adjusted R^2	0.185	0.146	0.192	0.138

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Regressions with individual and village forecast revisions.

	$(\frac{y_t}{e_{t,t-1}} - 1)^2$ - $\frac{\sigma^2_{t-1,t}}{e^2_{t-1,t}}$	$(pit_t - 0.5)^2 - \frac{1}{12}$	$(\frac{y_t}{e^f_{t,t-1}} - 1)^2$ - $\frac{\sigma^{2,f}_{t-1,t}}{e^{2,f}_{t-1,t}}$	$(pit^f_t - 0.5)^2 - \frac{1}{12}$
	(1)	(2)	(3)	(4)
forecast	−0.064	−0.025*	−0.199	−0.017
	(0.464)	(0.014)	(0.447)	(0.020)
village revision	0.828	0.022*	1.217*	0.026**
	(0.568)	(0.012)	(0.621)	(0.013)
individual revision	−0.642	−0.023	−0.876	−0.035
	(0.557)	(0.024)	(0.549)	(0.027)
Constant	−0.023	−0.006*	−0.010	−0.002
	(0.038)	(0.003)	(0.034)	(0.003)
Rationality (pval)	$< 10^{-12}$	0.011	1.4e-09	0.28
Observations	2,778	2,778	2,778	2,778
Adjusted R^2	0.145	0.111	0.174	0.100

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Regressions with individual and village forecast revisions.

4 Consumption Choice

In empirical studies consumption is commonly smoother than income. The permanent income hypothesis argues that mainly permanent, not transitory, income changes drive consumption (Hall, 1978). Two central mechanisms that allow economic agents to smooth consumption are insurance against changing income and advanced information about changing income. A central task of applied work on consumption is to disentangle the two channels. As the information available to the economic agent is generally not observable by the econometrician, even under rational information it is unclear how to construct subjective probabilities. A problem also referred to as *superior information issue* (Flavin, 1993). The literature provides several attempts to separate the two mechanisms with the use of subjective expectation data. Early work in this direction includes (Hayashi, 1985). In the following we apply several different approaches (see Jappelli and Pistaferri, 2010, for a review).

In the next section, we start with a naive log-linearized Euler equation that can be estimated directly from the data. In Section 4.2, we consider subjective expectations to separate permanent and transitory income shocks as implemented in Attanasio et al. (2020) and Pistaferri (2001). Based on this computation, one can estimate a log linearized Euler equation directly. In Section 4.3 we use expectations as instruments in a first stage regression for expected income to estimate a log linearized Euler equation accounting for irrationality and measurement error in expectations in an in a two-stage IV estimation. This approach was pioneered in Jappelli and Pistaferri (2000). We add the second moment of income to investigate precautionary savings in the presence of measurement error. In Section 4.4, we analytically derive the moments between consumption, income and expectations and estimate the underlying parameters in the model in a moment-based approach as first proposed in Kaufmann and Pistaferri (2009).

4.1 Naive Approach

We begin by analyzing simple associations of consumption changes to contrast the findings to those of the more sophisticated approaches that are to follow.

In the following we differentiate between the income in a expectation target month y_t^e and the income y_t^c in the 30 days that consumption was elicited for. As a reminder, we denote by c_t the consumption reported at time t, where the target is given by the past 30 days. y_t^c denotes the income in those 30 days. $e_{t,t+1}$ denotes the expectation elicited at time t (at the same time as the consumption). y_{t+1}^e denotes the income in the target month of expectation $e_{t,t+1}$. The days of overlap between the monthly income y_t^c and y_t^e depends on the timing of the interview and ranges from 0 to 30 days.

In Table 14 we explain log consumption changes. Column (1) shows that both the innovation in income as well as the innovation in expectations increase consumption. Income changes are further analyzed in column (2) to (7) by its two components

$$\Delta y_t^c = (y_t^c - e_{t-1,t}) + (e_{t-1,t} - y_{t-1}^c),$$

the expected income change $e_{t-1,t} - y_{t-1}^c$ and the unexpected income change $y_t^c - e_{t-1,t}$. Similar estimates emerge in this case in column (2). The results are robust to including household fixed effects in column (4), but unexpected income and change in expectations is found to influence consumption stronger for better performing forecasters in column (3).

A differentiation between positive and negative shocks paints a more heterogeneous picture of expectations, income and consumption: Column (5) to (7) provide evidence that expected income increases lead to much smaller consumption responses than expected decreases in income. In fact, in most specifications there is no evidence for a consumption response to expected income increases. On the contrary, unexpected income increases induce a strong consumption response. Here, unexpected income decreases do induce income changes that are statistically significant, but about 70% smaller. For changes in expectations no column suggests strong evidence for a heterogeneous response.

Under the permanent income hypothesis without saving constraints expected changes in income should not influence consumption. The positive coefficient for expected decreases in income suggests that households are unable to save before the expected income change or decide against consuming wealth in times of low income. A lack of saving is consistent with the large coefficient of unexpected (and presumably transitory) increases in income, which suggests that only a small part of this additional income is saved.

4.2 Computing Transitory and Permanent Shocks

The analysis of consumption response above can be refined for more precise and robust conclusions. Theoretical arguments connect consumption changes to changes in expected life cycle income. The change in expectations proxy for expected life cycle income, but the exact connection depends on assumptions about the underlying income process even under full information rational expectations.

Let us denote life cycle income by $\sum_{t'} y_{t'}$ and its expected value at time t by

$$\mathbb{E}_t[\sum_{t'} y_{t'}] = \sum_{t'} e_{t,t'}.$$

The change in expected life cycle income between t and $t+1$ is thus given by

$$\sum_{t'} e_{t,t'} - e_{t-1,t'}.$$

Our subjective probability data measures only the term $e_{t-1,t}$, $e_{t,t+1}$ and $e_{t,t} = y_t$, which leaves all but the difference $e_{t,t} - e_{t-1,t} = y_t - e_{t-1,t}$ to be determined by assumptions.

The change in expectations $\Delta e_{t,t+1} = e_{t,t+1} - e_{t-1,t}$ is another term in the expected life cycle income. The naive approach in Section 4.1 identifies the reaction to changes in life cycle income, if we assume that for all $t' > t$ it holds that $\Delta e_{t,t'} = \Delta e_{t,t+1}$ (i.e., the agent expects all future income to change the same amount as the next period income.)

Without expectation data it is hard to know if an income shock is transitory or permanent. Subjective expectations allow to identify transitory and permanent income shocks under specific data generating assumptions as outlined in Pistaferri (2001) and Attanasio et al. (2020). Following the aforementioned literature, we compute transitory and permanent income shocks by the difference in expectation after removing predictable lifecycle effects. In the context considered here the most important predictable income change is seasonal variation. As argued before, the fishing

	Δc_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δy_t^c	0.122**						
	(0.059)						
$e_{t-1,t} - y_{t-1}^c$ +					−0.304**	−0.259*	−0.305*
					(0.145)	(0.144)	(0.164)
$e_{t-1,t} - y_{t-1}^c$		0.127*	0.121*	0.133*	0.395**	0.341**	0.404**
		(0.068)	(0.067)	(0.074)	(0.171)	(0.151)	(0.190)
$y_t^c - e_{t-1,t}$ +					0.317*	0.337	0.325*
					(0.190)	(0.225)	(0.195)
$y_t^c - e_{t-1,t}$		0.116*	0.180**	0.111	0.081	0.127*	0.078
		(0.061)	(0.082)	(0.067)	(0.062)	(0.071)	(0.068)
$\Delta e_{t,t+1}$ +					0.072	0.065	0.128
					(0.135)	(0.358)	(0.197)
$\Delta e_{t,t+1}$	0.122	0.126	0.224	0.131	0.085	0.172	0.058
	(0.082)	(0.086)	(0.154)	(0.090)	(0.094)	(0.213)	(0.132)
comments	-	-	good	FE	-	good	hh FE
Observations	2,567	2,567	1,239	2,567	2,567	1,239	2,567
R^2	0.100	0.100	0.141	0.107	0.127	0.166	0.135
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 14: Naive regression for consumption changes. The dependent variable is innovations in log consumption. Standard errors are clustered at the time level. Columns commented with "good" are estimated on the better performing half of forecasters.

income is subject to substantial seasonal variation and changing expectations in line with typical seasonal variation do not necessarily imply a change in permanent income. Consequently, we define the permanent income shock (*pis*) Ψ as the difference between the change in expectations and the typical seasonal change at season s_t

$$\Psi_t = \Delta e_{t,t+1} - \Delta s_{t+1},$$

where s_t is defined as the average income in this month for a particular boat owner.

The temporary income shock (*tis*) ϵ can be identified by the difference between seasonally adjusted current income and currently expected income

$$\epsilon_t = (y_t - s_t) - (e_{t,t+1} - s_{t+1}).$$

Both computations are intuitively accessible, but can also be derived in seasonal model of income with permanent and transitory income shocks. Assume that income has the following structure

$$y_t = p_t + s_t + \epsilon_t \tag{1}$$

with the permanent income being

$$p_t = p_{t-1} + \Psi_t, \tag{2}$$

Under rational expectations, with the seasonal pattern s_t being known to the agent, it follows that $e_{t-1,t} = p_{t-1} + s_t$ and thus

$$\Delta e_{t,t+1} = \Delta p_t + \Delta s_{t+1}.$$

The formula for the permanent income shock $\Psi_t = \Delta p_t$ follows immediately. With Equation (1) the temporary income shock can be computed as

$$\begin{align}
\epsilon_t &= y_t - (p_t + s_t) \tag{3}\\
&= y_t - (p_t + s_{t+1} - s_{t+1} + s_t) \tag{4}\\
&= y_t - (e_{t,t+1} - \Delta s_{t+1}) \tag{5}\\
&= (y_t - s_t) - (e_{t,t+1} - s_{t+1}). \tag{6}
\end{align}$$

Compare Pistaferri (2001) and Attanasio et al. (2020) for similar computations without seasonal adjustments.

Figure 3 shows density estimates and a scatterplot for the computed permanent and transitory income shocks. We observe a strong negative correlation. Besides the elementary conclusion that temporary changes are connected to contrary long-term implications, there exist several alternative explanations: Measurement error in expectations, noisy expectation formation, and error added by subtraction of an estimated seasonal effect. A similar correlation coefficient is observed for the subset of better forecasting boat owners, which suggest that the correlation is not (entirely) driven by measurement error. Possibly, long-term income increases are bought with temporary reduced income and long-term income reductions are compensated temporarily by increasing income (e.g., going fishing more often).

We observe am empirical variance for the permanent shocks of 0.04 and for the transitory shocks of 0.06, which is considerably higher than the findings in Blundell et al. (2008), who observe

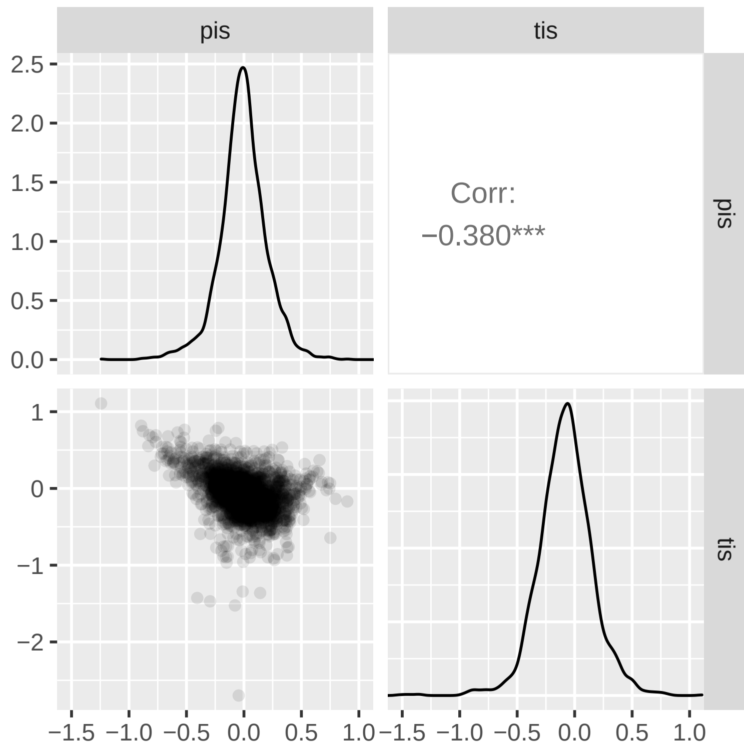

Figure 3: Scatterplot, histograms, and sample Pearson correlation coefficient for permanent and temporary income shocks.

a variance for permanent shocks of 0.01 to 0.03 and for transitory shocks of 0.02 to 0.05 for annual data.

Table 15 shows regressions of the shocks on log consumption changes. The results suggest that permanent income shocks influence consumption to a higher extend than transitory shocks, as predicted by the permanent income hypothesis. As shown in columns (3) to (7), the non-zero coefficient of transitory income shocks is explained mostly by negative shocks, which suggests borrowing constraints when faced with a negative impact shock.

The effect estimates are 2-3 times stronger in column (2) and (4) after dropping survey participants with subjective expectations that do not have high explanatory power. Those can be expected to deliver bad proxies for actual held beliefs. Such noisy measurements are known to bias effect estimates towards zero. Controlling with individual fixed effects in column (5), month fixed effects in column (6) or individual specific seasonal fixed effects in column (7) does not alter the conclusions, but reduces accuracy of point estimates.

4.3 Expectations as Instruments

We noted the strong evidence for measurement error in expectations, especially for the subjective variance. In the following we draw on Jappelli and Pistaferri (2000) and approximate expected income in a first stage regression by the subjective mean, which allows consistent estimates with measurement error in expectations at the cost of assuming rational expectations.

Generally, after modeling the conditional expectation of income y given a vector of covariates X^1, the log linearized Euler equation

$$\Delta \ln c_t = \beta_e E[\Delta y_t | X_t^1] + \beta X_t^2 + \epsilon_t,$$

can be estimated in a second step. The underlying assumption is that the statistically constructed expectation $E[y|X^1]$ coincides with the subjective expectations. This assumption implies rational expectations and that the covariates X^1 indeed reconstruct the information set of the agent.

The innovation in Jappelli and Pistaferri (2000) lies in using elicited expectations in the first stage regression as a covariate. Conveniently, measurement error in the subjective probabilities would not invalidate this estimation procedure. The elicited subjective probabilities can be argued to be better predictors of the conditional expectation of income in the first stage. The subjective probabilities can be more specific to the individual than variables commonly available to the econometrician in such settings (e.g, past income, age, number of hh members, occupation, etc.). In fact, the approach to construct $E[y|\mathcal{F}]$ by regressing such variables on income is bound to fall short if the information set of the economic agent is larger (referred to as *superior information issue* in Flavin, 1993). In the consumption income setting this can be expected, as a whole set of potential information could be available to the individual beforehand (e.g., promotions, health issues, layoffs, education choice, caregiving, etc.). Importantly, even if such information is encoded in available covariates when realizing, this would not capture the timing of when such an event was incorporated in the information set, nor the potential continuous transition of the perceived likelihood of such an event. Subjective expectations could, in theory, capture all the information relevant for an outcome.

Considering the evidence for rational inattention (Sims, 2003) and bounded rationality (Kahneman, 2003), it is also possible that some information available to the econometrician (e.g., past

	Δc_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
pis_t	0.134*	0.297**	0.107*	0.238**	0.114	0.146*	0.326
	(0.074)	(0.138)	(0.062)	(0.111)	(0.073)	(0.084)	(0.247)
pis_t +			0.040	0.107	0.055	0.030	0.084
			(0.083)	(0.246)	(0.107)	(0.082)	(0.237)
tis_t	0.060	0.110	0.078*	0.141**	0.112*	0.051	0.097
	(0.043)	(0.068)	(0.044)	(0.059)	(0.065)	(0.037)	(0.079)
tis_t +			−0.062	−0.094	−0.069	−0.032	−0.124
			(0.081)	(0.120)	(0.092)	(0.101)	(0.181)
good	-	Yes	-	Yes	-	-	-
FE	-	-	-	-	hh	month	hh × month
Observations	2,774	1,339	2,774	1,339	2,774	2,774	2,774
R^2	0.020	0.062	0.021	0.063	0.026	0.145	0.448
Note:					*p<0.1; **p<0.05; ***p<0.01		

Table 15: The effect of permanent and transitory income shocks. The dependent variable is innovations in log consumption. Standard errors are clustered at the time level. The comment "good" denotes regressions with the better performing half of forecasters.

income) is not used consistently to construct expectations (see also the evidence in Section 3.5). In this case, using the econometrician's information set, would make the constructed expectations overly sensitive to changes in the variables. Using the subjective probabilities in the first stage regression would induce the preferable property of making the constructed expectations insensitive to the information not included in the expectation formation process.

While the approach imposes few restrictions on the validity and quality of subjective expectation data, it does assume that consumption choices are based on rational expectations as constructed in the first stage regression. One motivation behind the elicitation of subjective expectation data is the potential deviation from rationality, which could invalidate the two-stage approach here. Furthermore, equating expectations with post-hoc statistical models can be seen as not only imposing rationality, but some super rationality without learning that incorporates information available after the actual expectation formation took place.

We start with the first stage regression in Table 16, which regresses the first difference of realized log income in column (1) to (4) and realized squared error in column (5) to (8) on the first difference of mean and variance of the subjective distribution of log income. We use the raw data and seasonally adjusted data. The predictive power of the first stage regression for the mean in column (1) to (4) is high with an R^2 of about 0.5 without and 0.25 with seasonally adjusted income and expectations. The predictive content of the subjective probabilities for uncertainty in column (5) to (8) is much lower. While the variance coefficient is estimated to be distinct from zero in three of the four specifications, the R^2 value remains close to 0.01 indicating that almost all of the variation remains unexplained. Without seasonal adjustments, expectations are more predictive of income increases, whereas an expected decline does not hold the same predictive power. After seasonal adjustments no such evidence remains. For the uncertainty judgment in column (5) to (8) an increase in perceived uncertainty is not estimated to show any different behavior from a decrease irrespective of seasonal adjustments.

Second, we consider the two-stage IV regression by using the subjective means and variances as excluded instruments for predictable income changes, which avoids assuming rational expectations and relies instead on the assumption that subjective expectations are correlated with the conditional expectation of income and that the elicited subjective expectations influence consumption only through the expectation.

In Table 17, we consider predictable income changes. By using the change in consumption at time t and instruments that realized before, i.e. the subjective probabilities elicited at time $t-1$, the first stage regression constructs predictable income changes $\mathbb{E}[\Delta y_t|\mathcal{F}_{t-1}]$. Throughout, we control for actual income changes that are correlated with expectations. Without seasonal adjustments in column (1), we find no evidence for predictable income changes to influence consumption in line with the permanent income hypothesis, and we observe a negative coefficient for changes in predictable uncertainty, which is in line with precautionary savings. After seasonal adjustments in column (2), the effect of precautionary savings essentially vanishes and we observe a strong effect of predictable income increases (but none for decreases) in column (4). Columns (5) to (8) mirror those findings with smaller estimated coefficients without the use of a two-stage procedure, instead relying on the subjective self-reports directly. This difference suggests that measurement error substantially effects estimation.

One explanation for the different results between seasonally adjusted and the raw data could be that households are mostly able to save and borrow to smooth seasonal deviations from mean

	Δy_t^c				$\Delta(y_t^c - e_{t-1,t})^2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$e_{t-1,t} - y_{t-1}^c$ $^+$		0.877***		0.036				
		(0.187)		(0.385)				
$e_{t-1,t} - y_{t-1}^c$	0.967***	0.250	0.652***	0.632**	0.384	0.383	0.152	0.151
	(0.092)	(0.165)	(0.127)	(0.280)	(0.383)	(0.384)	(0.300)	(0.300)
$\Delta\sigma^2_{t-1,t}$ $^+$						−0.026		−0.078
						(0.151)		(0.156)
$\Delta\sigma^2_{t-1,t}$	−0.018***	−0.015**	−0.014**	−0.014**	0.147***	0.160	0.166***	0.205**
	(0.007)	(0.006)	(0.006)	(0.006)	(0.048)	(0.098)	(0.040)	(0.092)
seasonal adj.	no	no	yes	yes	no	no	yes	yes
Observations	2,576	2,576	2,576	2,576	2,367	2,367	2,367	2,367
R^2	0.514	0.545	0.246	0.246	0.008	0.008	0.010	0.011
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 16: First stage of IV regression for innovations of income and income uncertainty. Columns marked with "seasonal adj." all variables are seasonally adjusted by using the residuals from a regression on individual specific seasonal dummies.

income, but are unable to borrow money if expecting an income increase that is not seasonal and therefore unobservable for potential lenders.

In Table 18 we mirror the analysis above for predicted income changes by considering the connection between consumption changes and future income changes predictable with current instruments, i.e. instruments and consumption change are from the same time period $t-1$. This applications was not considered in Jappelli and Pistaferri, 2000. As outlined before, if predicted income changes proxy for changes in expected life cycle income, the respective coefficient is informative on the permanent income hypothesis. Similar to predictable income changes, we observe different responses to predicted income changes depending on seasonal adjustments. The average effect for the unadjusted data is lower in column (1), than the effect estimated after adjusting in column (2). Column (3) and (4) reveal that unadjusted changes (including seasonal variation) triggers a consumption response only for increased expectations and the adjusted data (controlling for seasonal change) only for negative changes. This suggests that households are able to save if expecting income declines in the future, but not able to borrow if expecting permanent income increases beyond the seasonal fluctuations. This interpretation is also supported by the results in column (7) and (8), where the direct usage of expectation data without a two-stage approach indicates lower point estimates and standard errors.

Changes in perceived uncertainty show no evidence for an effect on consumption. The two-stage approach in column (1) to (4) leads to a 10 to 20 times higher point estimate, which remains statistically insignificant in all specifications. Either precautionary saving is not relevant in long-term uncertainty, or the short-term expectations are not a sufficiently sharp proxy for long-term uncertainty.

	Δc_t							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δy^c_{t-1}	−0.130	−0.013	−0.142	−0.031	−0.131*	−0.084	−0.131*	−0.083
	(0.081)	(0.089)	(0.092)	(0.094)	(0.077)	(0.098)	(0.077)	(0.095)
Δy^c_t +			0.229	0.631**				
			(0.381)	(0.287)				
Δy^c_t	0.011	0.289*	−0.175	−0.090				
	(0.043)	(0.150)	(0.318)	(0.164)				
$\Delta(y^c_t - e_{t-1,t})^2$	−0.045	−0.005	−0.050	−0.013				
	(0.030)	(0.023)	(0.036)	(0.025)				
$e_{t-1,t} - y^c_{t-1}$ +							−0.001	0.396*
							(0.073)	(0.208)
$e_{t-1,t} - y^c_{t-1}$					−0.001	0.149**	0.0003	−0.086
					(0.040)	(0.067)	(0.062)	(0.104)
$\Delta\sigma^2_{t-1,t}$					−0.008**	−0.005	−0.008**	−0.005
					(0.004)	(0.004)	(0.004)	(0.004)
seasonal adj.	no	yes	no	yes	no	yes	no	yes
IV	yes	yes	yes	yes	no	no	no	no
Observations	2,325	2,325	2,325	2,325	2,535	2,535	2,535	2,535
R^2	0.005	0.059	0.001	0.045	0.092	0.058	0.092	0.076
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 17: Second stage of IV regression for predictable income change. The dependent variable is consumption change. Columns marked with "seasonal adj." all variables are seasonally adjusted by using the residuals from a regression on individual specific seasonal dummies. See Table 16 for the excluded instruments.

	Δc_{t-1}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δy^c_{t-1}	0.180*	0.320	0.162	0.331	0.153*	0.222	0.162*	0.222
	(0.092)	(0.213)	(0.114)	(0.202)	(0.084)	(0.159)	(0.084)	(0.155)
Δy^c_t +			0.326	−0.426				
			(0.250)	(0.394)				
Δy^c_t	0.109	0.316	−0.154	0.573				
	(0.069)	(0.266)	(0.243)	(0.396)				
$\Delta(y^c_t - e_{t-1,t})^2$	0.021	0.039	0.013	0.045				
	(0.020)	(0.030)	(0.022)	(0.033)				
$e_{t-1,t} - y^c_{t-1}$ +							0.206***	−0.316
							(0.079)	(0.199)
$e_{t-1,t} - y^c_{t-1}$					0.079	0.158	−0.083	0.339*
					(0.061)	(0.133)	(0.063)	(0.200)
$\Delta\sigma^2_{t-1,t}$					0.002	0.002	0.002	0.002
					(0.002)	(0.004)	(0.002)	(0.004)
seasonal adj.	no	yes	no	yes	no	yes	no	yes
IV	yes	yes	yes	yes	no	no	no	no
Observations	2,324	2,324	2,324	2,324	2,740	2,740	2,740	2,740
R^2	0.029	0.022	0.041	0.027	0.081	0.061	0.089	0.073
Note:						*p<0.1; **p<0.05; ***p<0.01		

Table 18: Second stage of IV regression for predicted income change. The dependent variable is lagged consumption change. Columns marked with "seasonal adj." all variables are seasonally adjusted by using the residuals from a regression on individual specific seasonal dummies. See Table 16 for the excluded instruments.

4.4 Moment-Based Approach

Kaufmann and Pistaferri (2009) propose to disentangle expectation dynamics and partial insurance by moment conditions. Income, expectation, and consumption are related by a set of assumptions, which imply parameter dependent moments between the three observed time series. Conveniently, the moment-based estimation procedure in the spirit of GMM estimation (Hansen, 1982) does not require a fully specified model and can, at least in principle, account for measurement error in all variables.

4.4.1 Model

The income model in Section 4.2 has been extended to include fixed effects in Guvenen (2007) and Storesletten et al. (2004). To account for seasonal deviation, we consider another slightly different model of the standard model used in Kaufmann and Pistaferri (2009) with lag differences of order 6 which amounts to yearly differences for the bimonthly panel data. The income process is denoted by

$$y_t = p_t + \epsilon_t, \tag{7}$$

with a transitory shock ϵ_t and a permanent component p_t that is generated with

$$p_t = p_{t-6} + \Psi_t, \tag{8}$$

where the permanent shock is denoted by Ψ_t. Note that we deviate here from the model most commonly used with $p_t = p_{t-1} + \Psi_t$. As the income from fishing is subject to strong seasonal patterns, a season-specific income process seems more realistic for our households. Similar arguments could be put forward for farming income.

Both shocks can be analyzed as an anticipated part (known to the agent at all times) and unanticipated part (learned at time t):

$$\epsilon_t = \epsilon_t^A + \epsilon_t^U, \quad \Psi_t = \Psi_t^A + \Psi_t^U.$$

Note that this model, while accounting for changing expectations about the income y_t, fails to accommodate changes in expectations about the shocks realizing at time t. To complete the model, we define the information set of the agent as

$$\mathcal{F}_t = \sigma(y_t, \epsilon_t, \Psi_t, \Psi^A, \epsilon^A),$$

where $\Psi^A = (\dots, \Psi_t^A, \Psi_{t+1}^A, \dots)$ and $\epsilon^A = (\dots, \epsilon_t^A, \epsilon_{t+1}^A, \dots)$ denote the vector of anticipated shocks at all time points t.

Consumption changes are assumed to react to unanticipated income changes only

$$\Delta^6 c_t = \beta_\epsilon \epsilon_t^U + \beta_\Psi \Psi_t^U. \tag{9}$$

From the income process it follows that

$$\Delta^6 y_t = \Delta^6 \epsilon_t + \Psi_t \tag{10}$$

and rational expectations imply $e_{t,t+1} = p_{t-6} + \Psi_t^A + \epsilon_t^A$ and thus

$$\Delta^6 e_{t,t+1} = \Psi_{t-6} + \Delta^6 \Psi_t^A + \Delta^6 \epsilon_t^A. \tag{11}$$

The left-hand side of equations (9) to (11) is observed and can be used to construct moment conditions. We include measurement error in expectations r_e and consumption r_c. Again following Kaufmann and Pistaferri (2009), we use the following moments. With income we have

$$\mathbb{E}[(\Delta^6 y_t)^2] = 2\sigma_\epsilon^2 + \sigma_\Psi^2, \tag{12}$$
$$\mathbb{E}[\Delta^6 y_t \Delta^6 y_{t-6}] = -\sigma_\epsilon^2. \tag{13}$$

With the expectations we have

$$\mathbb{E}[(\Delta^6 e_t)^2] = \sigma_\Psi^2 + 2\sigma_{\epsilon^A}^2, \tag{14}$$
$$\mathbb{E}[\Delta^6 e_t \Delta^6 e_{t-6}] = -\sigma_{\epsilon^A}^2 - \sigma_{r_e}^2. \tag{15}$$

With consumption we have

$$\mathbb{E}[(\Delta^6 c_t)^2] = \beta_\Psi^2 \sigma_{\Psi^U}^2 + \beta_\epsilon^2 \sigma_{\epsilon^U}^2 + 2\sigma_{r_c}^2, \tag{16}$$
$$\mathbb{E}[\Delta^6 c_t \Delta^6 c_{t-6}] = -\sigma_{r_c}^2. \tag{17}$$

Additionally, we include the following five interactions

$$\mathbb{E}[\Delta^6 e_t \Delta^6 y_t] = 2\sigma_{\epsilon^A}^2 + \sigma_{\Psi^A}^2 \tag{18}$$
$$\mathbb{E}[\Delta^6 e_t \Delta^6 y_{t-6}] = -\sigma_{\epsilon^A}^2 + \sigma_{\Psi^U}^2 \tag{19}$$
$$\mathbb{E}[\Delta^6 e_t \Delta^6 c_{t-6}] = \beta_\Psi \sigma_{\Psi^U}^2 \tag{20}$$
$$\mathbb{E}[\Delta^6 c_t \Delta^6 y_t] = \beta_\Psi \sigma_{\Psi^U}^2 + \beta_\epsilon \sigma_{\epsilon^U}^2 \tag{21}$$
$$\mathbb{E}[\Delta^6 c_{t-6} \Delta^6 y_t] = -\beta_\epsilon \sigma_{\epsilon^U}^2 \tag{22}$$
$$\tag{23}$$

In total, we observe the eleven empirical moment conditions above and seek to estimate eight parameters (the tendency to consume $\beta_\Psi, \beta_\epsilon$, the variance of the different shocks $\sigma_{\Psi^A}^2, \sigma_{\Psi^U}^2, \sigma_{\epsilon^A}^2, \sigma_{\epsilon^U}^2$, and the measurement errors $\sigma_{r_c}^2, \sigma_{r_e}^2$).

4.4.2 Estimation

Kaufmann and Pistaferri (2009) follow other papers in the consumption smoothing literature (e.g, the seminal paper Blundell et al., 2008) and use an Equally Weighted Minimum Distance (EWMD), as proposed in Altonji and Segal (1996), which is a GMM estimator with a particular simple weighting matrix.

Altonji and Segal (1996) argue that the asymptotically vanishing bias induced by the covariance estimation for efficient GMM can lead to decreased efficiency in finite samples compared to the simple EWMD. It is however worth noting that sample sizes in the consumption response setting are often much larger than the finite sample size considered in the simulations in Altonji and Segal (1996). Further, the continuously updated GMM estimator was found to be less biased (Hansen et al., 1996) than standard two-step GMM, and Clark (1996) finds similar biases in the overidentifying restriction test for EWMD and efficient GMM. Overall, EWMD seems not a particular convincing solution to the challenging estimation problem for the moments derived in Section 4.4.1.

An additional concern is weak or partial identification, which may invalidate any of the approaches considered above. The same applies to indirect inference as for example used by Guvenen and Smith (2014) to disentangle insurance and information.

Instead, we propose to rely on Bayesian GMM[2] as proposed in Chernozhukov and Hong (2003) and Chen et al. (2018).

We begin by introducing standard GMM before providing some intuition for the Bayesian GMM. Given a set of parametric moment conditions

$$E[g(X,\theta)] = 0$$

for some q-dimensional function g and some p dimensional paramter θ, the GMM estimator is given by the minimizer of some quadratic norm of the moment function g, i.e.

$$\hat{\theta}_{GMM} = \arg\min_{\theta} \, g(X,\theta)' W g(X,\theta),$$

for some weighting matrix W. An efficient choice for W is the inverse of the covariance matrix of g. As the covariance matrix is unknown and can be estimated only given a valid parameter, one has to rely on multistep procedures like two-step, iterative, or simultaneously updated GMM.

The economic literature often refers to GMM where the weighting matrix is a consistent covariance estimator as OMD. The EWMD arises after setting non-diagonal elements to zero (ignoring any correlation between the moments).

To provide consistent inference, the approaches above require the parameter to be identified in the sense that there exists only one element θ_0 in the parameter space such that $E[g(X,\theta_0)] = 0$. Interestingly, this assumption is not necessary for inference and a partial identification approach can be used instead. As shown in Stock and Wright (2000) the target $g(X,\theta)'Wg(X,\theta)$ follows a Chi-squared distribution at any parameter with $E[g(X,\theta)] = 0$ if W is a consistent covariance estimator for the given parameter value (as used in the continuously updated GMM). This can be used to construct asymptotically valid confidence sets under weak or partial identification (Stock and Wright, 2000). The procedure requires computation of the target $g(X,\theta)'Wg(X,\theta)$ over a dense grid in the parameter space, which becomes computationally unfeasible for high dimensions. A computationally attractive alternative can be found in MCMC approaches as first proposed in Chernozhukov and Hong (2003), where transition probabilities are governed by

$$exp(-g(X,\theta)'Wg(X,\theta)). \tag{24}$$

A relatively accessible intuition arises in comparison to standard Bayesian MCMC, where the posterior emerges as the product of likelihood and prior. Without a fully specified model, the

[2]The term Bayesian GMM is taken from Yin (2009). The method is also referred to as stochastic GMM or Laplace type estimators, where the latter is more general but coincides with the methods presented here if the objective function is moment-based.

likelihood of the observed data is not well-defined. However, given the moment conditions we can infer that our empirical moments are asymptotically normal by invoking a central limit theorem. Thus, if the weighting matrix W approximates the inverse of the covariance matrix, Equation (24) can be understood as the *quasi-posterior* of the data given a parameter θ using the *approximate likelihood* of a central limit theorem. Chernozhukov and Hong (2003) and Chen et al. (2018) provide sufficient conditions for Bernstein-von-Mises Type results that establish that indeed the quasi-posteriors based on the approximate likelihood provide consistent frequentist inference.

We note that the same intuition lies behind earlier work on DSGE models in Christiano et al. (2010). Noteworthy, Christiano et al. (2010) also rely on a diagonal covariance estimate basically transferring the EWMD to the quasi Bayes setting. It should however be noted that the GMM objective function without a consistent estimator of the covariance does not fulfill the Generalized Information Equality condition of Chernozhukov and Hong (2003). Thus, the posterior cannot be used directly to compute the variance of the estimates and requires additional steps before valid inference statements can be made.

4.4.3 Results

We applied the Bayesian GMM with uninformative priors[3]. As weighting matrices we used the inverse of an Heteroscedasticity-consistent (HC) covariance estimator and the diagonal version (diag) in analogy to the EWMD estimator setting correlations to zero. We consider the full sample, the good forecaster and the best forecasters (defined by the average ratio of explained variation in income.)

Graphical convergence diagnostics and Rhat statistics below 1.02 provide no evidence against the MCMC having converged to its stationary distribution.[4]

The quasi-posteriors obtained for all combinations are shown in Figure 4. We observe a small tendency to consume from transitory shocks β_ϵ from 0.02 to 0.10 and a much larger ratio of 0.30 to 0.75 for permanent shocks β_Ψ. Importantly, the uncertainty for the propensity to consume is considerable with confidence intervals that are consistent with 0.25 to 0.90.

Good forecasters adapt less to permanent shocks relative to the shock, but shocks are also estimated to be about twice as large. Transitory unexpected shocks are much larger than permanent unexpected shocks. Anticipated shocks are much smaller than unexpected shocks in general. Across all specifications the mean posterior of the transitory unexpected shock is largest with 0.12, followed by the permanent unexpected shock with 0.05. The anticipated shocks both have smaller posterior means of 0.01.

We estimate a considerable measurement error for consumption suggesting that simple regression approaches with first differences of log consumption are subject to biases. Expectations also exhibit large measurement error, less so for the good forecasters.

The diagonal covariance matrix (parallel to EWMD estimator) provides sharper uncertainty statements for propensity to consume temporary income shocks, but not for the variances of the income process or the measurement error.

[3]A uniform prior on the unit interval for the propensity to consume β_Ψ and β_ϵ and a normal distribution with standard deviation 5 truncated at zero for the remaining standard deviations.

[4]We note, however, that including clustered covariance estimators in results not shown here did make convergence more challenging.

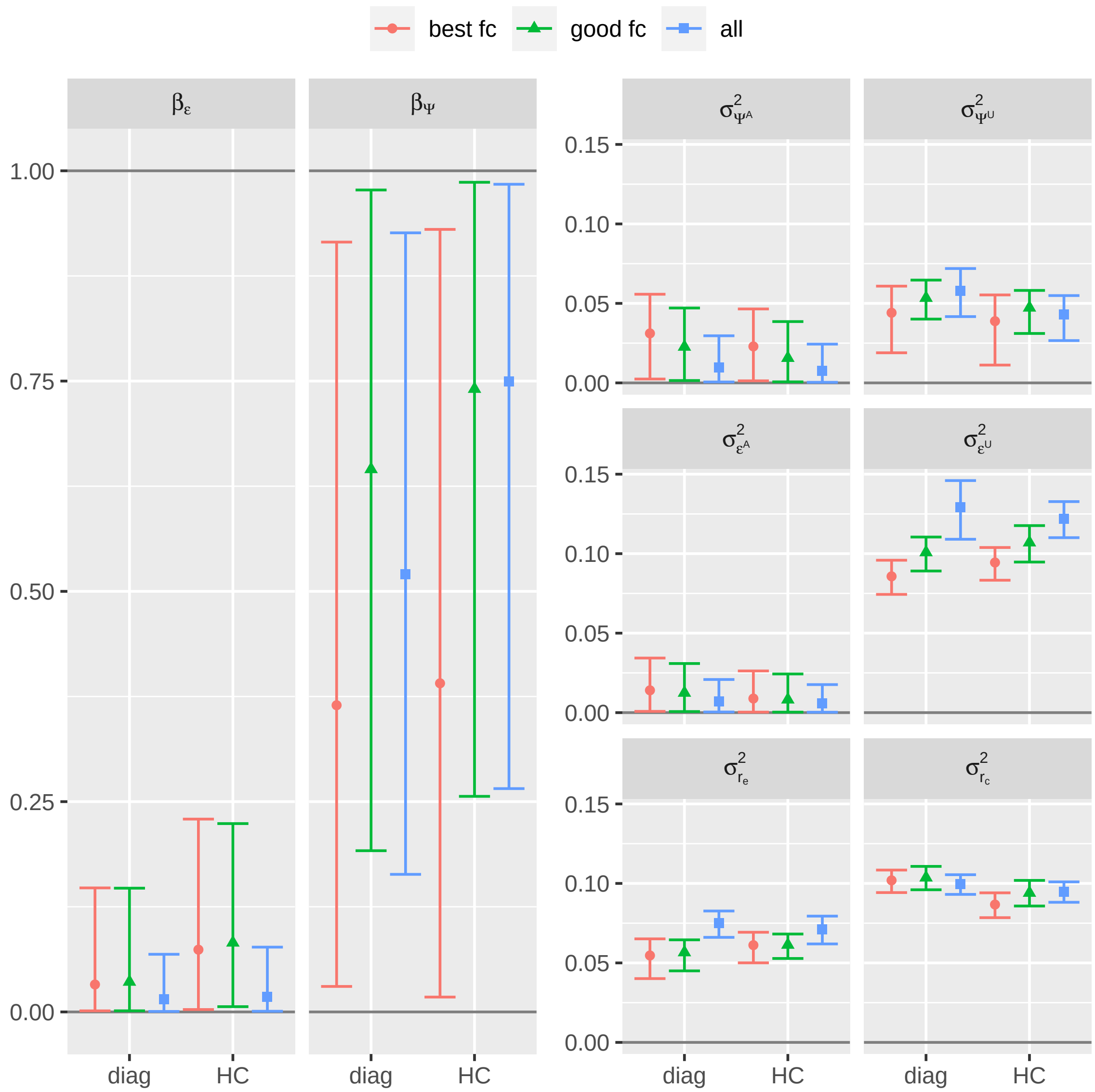

Figure 4: Quasi-posterior result of Bayesian GMM for moment-based approach. Median and 95% confidence intervals are shown. The left panel shows the propensity to consume from transitory shocks β_ϵ and from permanent shocks β_Ψ. The right panel shows the variances of the anticipated permanent shock $\sigma^2_{\Psi^A}$, the unanticipated permanent shock $\sigma^2_{\Psi^U}$, the anticipated temporary shock $\sigma^2_{\epsilon^A}$, the unanticipated temporary shock $\sigma^2_{\epsilon^U}$, and the measurement errors for expectations $\sigma^2_{r_e}$ and consumption $\sigma^2_{r_c}$.

5 Conclusion

The literature on the use of subjective probabilities in economic modeling generally concludes that belief measurement is feasible and informative (Manski, 2004; Manski, 2018; Delavande et al., 2011b; Attanasio, 2009). We note that the empirical evidence is based almost entirely on forecasts of central tendency (for example the mean of a subjective distribution, the mean elicited directly, or some unspecified point forecast). Extending forecast evaluation to uncertainty statements, we find that those are much less informative. We conclude that either respondents are less sensitive to changes in uncertainty or unable to communicate them in the used elicitation procedure. This calls for more research into new elicitation mechanisms and design improvements to elicit uncertainty perception (see for example Eyting and Schmidt, 2021).

We find that measurement error is a major concern for subjective probabilities, especially for uncertainty statements. Repeated measurements helped to alleviate some of the issues. We argue that using a second elicitation with a different household member constitutes a particular robust instrumental variable. Yet, possible channels for correlated measurement error remain. For example, if the household heads communicate about their responses (as opposed to their beliefs) or if the elicitor is aware of and influenced by the other household head's response. It should be noted that even if some correlated measurement error persists, the two-stage IV solution still reduces the bias compared to standard least squares by accounting for some of the measurement error (Gillen et al., 2019).

We reject full information rationality decidedly even after controlling for classical measurement error. One particular consistent deviation from rationality is the over-sensitive forecast, where high(low) forecasts are not followed by sufficiently high(low) realizations. A similar pattern of overstatement in anticipation of extreme events arose in Schmidt et al. (2021) for professional forecasts. In economic work, rational expectations might be defined differently. It is for example possible that over-sensitive beliefs improve decision making, when intersected with aversion to change.

Subjective probabilities can help disentangle different economic mechanisms (Manski, 2004). Here, we build on previous work to distinguish between information and insurance in consumption decisions (Kaufmann and Pistaferri, 2009). We note, however, that a single elicitation of subjective probabilities solves this identifiability issue only under strict assumptions. In particular, we follow the literature and assume that beliefs about a specific variable remain unchanged until realization, as the timing of change in expectations is not identified without eliciting time series of expectations in each survey wave. A more realistic model of reality might allow beliefs to change persistently.

In this work, we abstracted from non-linear models. Non-linear extensions of the canonical income process considered in Section 4.2 and changed to accommodate seasonal variation in Section 4.4 have been considered in the literature (Arellano et al., 2017; Arellano, 2014) and have been argued to be more appropriate (De Nardi et al., 2019). Interestingly, the approach relies on quantile estimation in an intermediate step that could be substituted by the pit using subjective probability distributions. We leave this aspect of non-linear dynamics for future research.

References

Altonji, J. G. and L. M. Segal (1996). "Small-sample bias in GMM estimation of covariance structures". In: *Journal of Business & Economic Statistics* 14.3, pp. 353–366.

Arellano, M. (Oct. 2014). "Uncertainty, persistence, and heterogeneity: A panel data perspective". In: *Journal of the European Economic Association* 12.5, pp. 1127–1153.

Arellano, M., R. Blundell, and S. Bonhomme (2017). "Earnings and consumption dynamics: A nonlinear panel data framework". In: *Econometrica* 85.3, pp. 693–734.

Attanasio, O. and B. Augsburg (2016). "Subjective expectations and income processes in rural India". In: *Economica* 83.331, pp. 416–442.

Attanasio, O., B. Augsburg, and R. De Haas (2019). "Microcredit contracts, risk diversification and loan take-up". In: *Journal of the European Economic Association* 17.6, pp. 1797–1842.

Attanasio, O., A. Kovacs, and K. Molnar (2020). "Euler equations, subjective expectations and income shocks". In: *Economica* 87.346, pp. 406–441.

Attanasio, O. P. (2009). "Expectations and perceptions in developing countries: Their measurement and their use". In: *American Economic Review* 99.2, pp. 87–92.

Attanasio, O. P. and K. M. Kaufmann (2014). "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender". In: *Journal of Development Economics* 109, pp. 203–216.

Blundell, R., L. Pistaferri, and I. Preston (2008). "Consumption inequality and partial insurance". In: *American Economic Review* 98.5, pp. 1887–1921.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020). "Overreaction in macroeconomic expectations". In: *American Economic Review* 110.9, pp. 2748–82.

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar (2019). "Sticky expectations and the profitability anomaly". In: *The Journal of Finance* 74.2, pp. 639–674.

Browning, M. and A. Lusardi (1996). "Household saving: Micro theories and micro facts". In: *Journal of Economic Literature* 34.4, pp. 1797–1855.

Bunn, P., J. Le Roux, K. Reinold, and P. Surico (2018). "The consumption response to positive and negative income shocks". In: *Journal of Monetary Economics* 96, pp. 1–15.

Chen, X., T. M. Christensen, and E. Tamer (2018). "Monte Carlo confidence sets for identified sets". In: *Econometrica* 86.6, pp. 1965–2018.

Chernozhukov, V. and H. Hong (2003). "An MCMC approach to classical estimation". In: *Journal of Econometrics* 115.2, pp. 293–346.

Chong, Y. Y. and D. F. Hendry (1986). "Econometric evaluation of linear macro-economic models". In: *The Review of Economic Studies* 53.4, pp. 671–690.

Christiano, L. J., M. Eichenbaum, and D. Marshall (1991). "The permanent income hypothesis revisited". In: *Econometrica*, pp. 397–423.

Christiano, L. J., M. Trabandt, and K. Walentin (2010). "DSGE models for monetary policy analysis". In: *Handbook of Monetary Economics*. Vol. 3. Elsevier, pp. 285–367.

Clark, T. E. (1996). “Small-sample properties of estimators of nonlinear models of covariance structure”. In: *Journal of Business & Economic Statistics* 14.3, pp. 367–373.

Clements, M. P. and D. I. Harvey (2010). “Forecast encompassing tests and probability forecasts”. In: *Journal of Applied Econometrics* 25.6, pp. 1028–1062.

Coibion, O. and Y. Gorodnichenko (2012). “What can survey forecasts tell us about information rigidities?” In: *Journal of Political Economy* 120.1, pp. 116–159.

— (2015). “Information rigidity and the expectations formation process: A simple framework and new facts”. In: *American Economic Review* 105.8, pp. 2644–78.

Coibion, O., Y. Gorodnichenko, and S. Kumar (2018). “How do firms form their expectations? New survey evidence”. In: *American Economic Review* 108.9, pp. 2671–2713.

De Nardi, M., G. Fella, and G. Paz-Pardo (Mar. 2019). “Nonlinear household earnings dynamics, self-insurance, and welfare”. In: *Journal of the European Economic Association* 18.2, pp. 890–926.

De Nicola, F. and X. Giné (2014). “How accurate are recall data? Evidence from coastal India”. In: *Journal of Development Economics* 106, pp. 52–65.

Delavande, A. (2014). “Probabilistic expectations in developing countries”. In: *Annu. Rev. Econ.* 6.1, pp. 1–20.

Delavande, A., X. Giné, and D. McKenzie (2011a). “Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design?” In: *Journal of Applied Econometrics* 26.3, pp. 479–497.

— (2011b). “Measuring subjective expectations in developing countries: A critical review and new evidence”. In: *Journal of Development Economics* 94.2, pp. 151–163.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). “Evaluating density forecasts with applications to financial risk management”. In: *International Economic Review* 39.4, pp. 863–883.

Dimitriadis, T., A. J. Patton, and P. Schmidt (2019). “Testing forecast rationality for measures of central tendency”. In: *Available at SSRN 3466442*.

Dominitz, J. (2001). “Estimation of income expectations models using expectations and realization data”. In: *Journal of Econometrics* 102.2, pp. 165–195.

Dominitz, J. and C. F. Manski (1997). “Using expectations data to study subjective income expectations”. In: *Journal of the American statistical Association* 92.439, pp. 855–867.

Engelberg, J., C. F. Manski, and J. Williams (2009). “Comparing the point predictions and subjective probability distributions of professional forecasters”. In: *Journal of Business & Economic Statistics* 27, pp. 30–41.

Erev, I. and E. Haruvy (2016). “10. Learning and the Economics of Small Decisions”. In: *The Handbook of Experimental Economics, Volume 2*. Ed. by J. H. Kagel and A. E. Roth. Princeton University Press, pp. 638–716.

Eyting, M. and P. Schmidt (2021). “Belief elicitation with multiple point predictions”. In: *European Economic Review* 135. Forthcoming.

Flavin, M. (1993). “The excess smoothness of consumption: identification and interpretation”. In: *The Review of Economic Studies* 60.3, pp. 651–666.

Flavin, M. A. (1981). "The adjustment of consumption to changing expectations about future income". In: *Journal of Political Economy* 89.5, pp. 974–1009.

Gillen, B., E. Snowberg, and L. Yariv (2019). "Experimenting with measurement error: Techniques with applications to the caltech cohort study". In: *Journal of Political Economy* 127.4, pp. 1826–1863.

Giné, X., M. Martinez-Bravo, and M. Vidal-Fernández (2017). "Are labor supply decisions consistent with neoclassical preferences? Evidence from Indian boat owners". In: *Journal of Economic Behavior & Organization* 142, pp. 331–347.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007). "Probabilistic forecasts, calibration and sharpness". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.2, pp. 243–268.

Gneiting, T. and M. Katzfuss (2014). "Probabilistic forecasting". In: *Annual Review of Statistics and Its Application* 1, pp. 125–151.

Gneiting, T. and R. Ranjan (2013). "Combining predictive distributions". In: *Electronic Journal of Statistics* 7, pp. 1747–1782.

Griliches, Z. and J. A. Hausman (1986). "Errors in variables in panel data". In: *Journal of Econometrics* 31.1, pp. 93–118.

Guvenen, F. (2007). "Learning your earning: Are labor income shocks really very persistent?" In: *American Economic Review* 97.3, pp. 687–712.

Guvenen, F. and A. A. Smith (2014). "Inferring labor income risk and partial insurance from economic choices". In: *Econometrica* 82.6, pp. 2085–2129.

Hall, R. E. (1978). "Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence". In: *Journal of Political Economy* 86.6, pp. 971–987.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". In: *Econometrica*, pp. 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). "Finite-sample properties of some alternative GMM estimators". In: *Journal of Business & Economic Statistics* 14.3, pp. 262–280.

Hausman, J. (2001). "Mismeasured variables in econometric analysis: problems from the right and problems from the left". In: *Journal of Economic Perspectives* 15.4, pp. 57–67.

Hayashi, F. (1985). "The permanent income hypothesis and consumption durability: analysis based on Japanese panel data". In: *The Quarterly Journal of Economics* 100.4, pp. 1083–1113.

Jappelli, T. and L. Pistaferri (2000). "Using subjective income expectations to test for excess sensitivity of consumption to predicted income growth". In: *European Economic Review* 44.2, pp. 337–358.

— (2010). "The consumption response to income changes". In: *Annu. Rev. Econ.* 2.1, pp. 479–506.

— (2020). "Reported MPC and unobserved heterogeneity". In: *American Economic Journal: Economic Policy* 12.4, pp. 275–97.

Jeong, J. and G. S. Maddala (1991). "Measurement errors and tests for rationality". In: *Journal of Business & Economic Statistics* 9.4, pp. 431–439.

Kahneman, D. (2003). “Maps of bounded rationality: Psychology for behavioral economics”. In: *American Economic Review* 93.5, pp. 1449–1475.
Kaufmann, K. and L. Pistaferri (2009). “Disentangling insurance and information in intertemporal consumption choices”. In: *American Economic Review* 99.2, pp. 387–92.
Kaufmann, K. M. (2014). “Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns”. In: *Quantitative Economics* 5.3, pp. 583–630.
Klepper, S. and E. E. Leamer (1984). “Consistent sets of estimates for regressions with errors in all variables”. In: *Econometrica*, pp. 163–183.
Knüppel, M. (2015). “Evaluating the calibration of multi-step-ahead density forecasts using raw moments”. In: *Journal of Business & Economic Statistics* 33.2, pp. 270–281.
Kohlhas, A. N. and A. Walther (2021). “Asymmetric attention”. In: *American Economic Review* 111.9, pp. 2879–2925.
Kueng, L. (June 2018). “Excess sensitivity of high-income consumers”. In: *The Quarterly Journal of Economics* 133.4, pp. 1693–1751.
Manski, C. F. (2004). “Measuring expectations”. In: *Econometrica* 72.5, pp. 1329–1376.
— (2018). “Survey measurement of probabilistic macroeconomic expectations: progress and promise”. In: *NBER Macroeconomics Annual* 32.1, pp. 411–471.
Mincer, J. A. and V. Zarnowitz (1969). “The evaluation of economic forecasts”. In: *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. NBER, pp. 3–46.
Moore, D. A. and P. J. Healy (2008). “The trouble with overconfidence.” In: *Psychological Review* 115.2, p. 502.
Patton, A. J. and A. Timmermann (2007). “Properties of optimal forecasts under asymmetric loss and nonlinearity”. In: *Journal of Econometrics* 140.2, pp. 884–918.
Pistaferri, L. (2001). “Superior information, income shocks, and the permanent income hypothesis”. In: *Review of Economics and Statistics* 83.3, pp. 465–476.
Rossi, B. and T. Sekhposyan (2019). “Alternative tests for correct specification of conditional predictive densities”. In: *Journal of Econometrics* 208.2, pp. 638–657.
Schmidt, P., M. Katzfuss, and T. Gneiting (2021). “Interpretation of point forecasts with unknown directive”. In: *Journal of Applied Econometrics* 36.6, pp. 728–743.
Sims, C. A. (2003). “Implications of rational inattention”. In: *Journal of Monetary Economics* 50.3, pp. 665–690.
Stock, J. H. and J. H. Wright (2000). “GMM with weak identification”. In: *Econometrica* 68.5, pp. 1055–1096.
Storesletten, K., C. I. Telmer, and A. Yaron (2004). “Consumption and risk sharing over the life cycle”. In: *Journal of Monetary Economics* 51.3, pp. 609–633.
Tarozzi, A., A. Mahajan, B. Blackburn, D. Kopf, L. Krishnan, and J. Yoong (2014). “Microloans, insecticide-treated bednets, and malaria: Evidence from a randomized controlled trial in Orissa, India”. In: *American Economic Review* 104.7, pp. 1909–41.

Thornton, R. L. (2012). “HIV testing, subjective beliefs and economic behavior”. In: *Journal of Development Economics* 99.2, pp. 300–313.

Wiswall, M. and B. Zafar (2015). “Determinants of college major choice: Identification using an information experiment”. In: *The Review of Economic Studies* 82.2, pp. 791–824.

— (2021). “Human capital investments and expectations about career and family”. In: *Journal of Political Economy* 129.5, pp. 1361–1424.

Yin, G. (2009). “Bayesian generalized method of moments”. In: *Bayesian Analysis* 4.2, pp. 191–207.