Subjective Expectations and Income Processes in Rural India*

Orazio Attanasio† and Britta Augsburg‡

May 2015

Abstract

This paper uses unique primary data to analyze and characterize the process that generates household income of poor households in rural India. We analyze and use data on individual subjective expectations elicited directly from the respondents of a household survey. We describe how the data was elicited and discuss its validity and to what degree we can trust that it reflects agents' believes about the future. We then use the responses to the subjective answers to the expectations questions and a parametric assumption to fit, for each household in the sample, a probability distribution for future income. We then use the moments we can compute from this distribution, together with data for actual current income, to specify and estimate a dynamic model of household income. We find that our households face a very persistent income process: we cannot reject the hypothesis of a random walk. Our paper is one of the first that uses subjective expectations data to model income processes.

*We are grateful to Manuel Arellano and Costas Meghir for some useful suggestions. We also benefited from useful comments from the seminar audience at the Research Department of the World Bank. This research was partly financed with funds from ESRC grants RES-000-22-3461, RES-167-25-0124 and RES-051-27-0135 and from ERC grant 249612 on "Exiting Long Run Poverty: The determinants of asset accumulation in developing countries" .

†University College London, EDePo@The Institute for Fiscal Studies, NBER, BREAD, o.attanasio@ucl.ac.uk. Tel: +44 (0)20 7679 5880.

‡EDePo@The Institute for Fiscal Studies, britta_a@ifs.org.uk. Tel: +44 (0)20 7291 4800. Fax: +444 (0)20 7323 4780

1 Introduction

Beliefs and expectations play a major role in decision-making processes. In particular, expectations about future income and, more generally, the nature of the stochastic process that generates income, determine, within models of intertemporal optimization, the allocation of current income between consumption and savings. Similarly, uncertainty about income prospects affects investment choices. Income expectations, however, are rarely observed and there is no consensus on the nature of income processes.

Most empirical studies infer expectations from data on actual income realizations appealing to some sort of rational expectations. Rational expectations assumptions are strong and, by and large, untested. Moreover, as one does not observe the information set available to agents, it is difficult to distinguish between individual heterogeneity and uncertainty, especially for researchers working with longitudinal data sets characterized by a large number of individuals (N) and relatively small number of periods (T).[1] And yet, in many contexts this distinction is very important as, for instance, in precautionary savings models.

A recent literature, partly surveyed by Manski (2004), has advocated the use of subjective expectations to tackle these issues and relax the assumption of rational expectations. Some advances have been made to assess the link between directly measured income expectations and realizations as well as other household characteristics. Specifically noteworthy are a series of papers by Dominitz and Manski (1996, 1997a,b) in which the authors develop and apply a useful framework for the measurement of subjective (income) distributions. These distinguish themselves from other contributions in this field by being able to estimate a respondent specific subjective expected income distribution, rather than having to work with point expectations as done in for example Das & van Soest (1997) and Alessie et al. (1997). A common finding in these papers is that realized household income is the

[1] An often cited example is Carroll (1994) who uses two different approaches to do so: in one, he assumes that people form their expectations about future income based on the realized income of older households with similar characteristics. In the second approach, he relates actual income over a period to personal characteristics and infers expectations accordingly.

dominant predictor of expected future household income - a finding also confirmed in our analysis.

With panel data sets available, Das & van Soest (1997) and Dominitz (1998) are able to go a step further and also compare the expectations to actual realizations over the same period. Finding that these match reasonably well but not entirely so, both studies argue that their results could be interpreted as evidence against the hypothesis of rational expectations.

In this paper, we follow this literature and construct and work with a subjective income distribution for each respondent household and compare the expectations to realizations over the same period. We will argue, however, that such a comparison, does not constitute a test of rational expectations. It is however, informative of the quality of the subjective expectations data.

The main contribution of our paper is instead the characterization of income processes and, most importantly, how they are perceived by agents. Ours is one of the first attempts to use subjective expectations data in such a way.[2] We see our approach as an alternative to use dynamic panel data methods (such as in MaCurdy, 1983, Abowd and Card, 1989, or Meghir and Pistaferri, 2004, to mention just a few studies) to characterize the stochastic properties of income processes.

Such a characterization of income processes is particularly important in developing countries, where income is much more volatile than in developed economies. A series of papers has analyzed stochastic dynamic income processes, in particular to examine the evidence of poverty traps (Jalan & Ravallion, 2001; Lokshin & Ravallion, 2004; Antman & McKenize, 2006).[3] To the best of our knowledge, this is however the first paper studying the features of income processes and what determines them incorporating the use of

[2] A similar approach is used by Attanasio and DiMaro (in progress, using data from rural Mexico). Unlike our sample, Attanasio and DiMaro's data do not have a longitudinal dimension, which somewhat limits the scope of their analysis.

[3] A strand of the literature looks similarly at asset dynamics, rather than income dynamics. Closest to the context of our study would be the study by Naschgold (2012), analyzing poverty dynamics in three villages in rural semi-arid India. Findings suggest "a strong type of poverty trap".

subjective expectation data in a developing country context.[4]

Of the few studies that have analyzed and used subjective income expectations, most concentrate on high-income countries. Our study distinguishes itself from these through the nature of our study population, who are poor households in rural India. Only recently a few surveys have started to collect data on subjective expectations in the developing world, where collection of such data might be particularly difficult because of the level of formal education of respondents and their lack of exposure to the formal concept of probability, a crucial input into the construction of expectation data. Attanasio (2009) discussed the progress made with respect to measurement of these variables. More recently, Delavande, Giné & McKenzie (2011) provide an overview of the recent contributions to this strand of literature and conclude, in line with Attanasio (2009), that eliciting subjective expectation data in developing countries is "feasible and valuable". Studies that have used subjective expectation data in a development setting are Attanasio et al (2005) on the use of probability distributions of future income in Colombia, and Luseno et al. (2003), Lybbert et al. (2007) on pastoralists' rainfall expectations in East Africa, and McKenzie et al. (2013) on income expectations of Tongans if they were to migrate to New Zealand.

The population we consider are households that depend on agriculture as their main source of income – as farmers or as agricultural laborers – while living in the second most drought prone area in India (Anantapur in the South of the state Andhra Pradesh). In this district, the average annual rainfall is not only extremely low but also highly variable and erratic. The district experiences prolonged dry spells of up to 50 days. As a result, during the years 1993 to 2006, there were only four 'good' years with better rainfall distribution during the cropping season and nine were 'drought' years. The interviewed households belong to the poorer part of the population within this area. These households were part of

[4] A very interesting paper by Santos & Barret (2006) studies wealth dynamics among poor pastoralists in rural Ethiopia, taking explicitly into account the role of individual heterogeneity as well as making use of pastoralists' expectations. They however concentrate on heard (i.e. asset) dynamics rather than income dynamics and use the expectation data to show that pastoralists perceive the non-linear long-term dynamics that characterize livestock wealth in the region.

a survey designed to evaluate a microfinance intervention. They were asked questions about their expectations on future income. In what follows we analyze these data to establish their plausibility and then use them to model the process that generates these households income and to estimate the parameters of this process.

More details on the data and respondent households are given in the next section, which also provides information on how the respondents' expectations were elicited in the course of the interview. Given the paucity of low-income country specific subjective income expectations, we then describe in some detail the elicited information and a validation of the same. Assuming a distributional form of the expectation data allows calculating moments of the expected outcome distributions in Section 3. In this section, we also analyze how the expectation data can be explained by among other variables, income realizations in the past. Section 4 proposes a simple statistical model for income. This model is estimated with our data in Section 5. We devote special attention to the way we model persistence and allow for fixed household effects. Section 6 concludes.

2 Background and Description of the Data

The data used in this study were collected as part of an evaluation effort of a micro-finance intervention in India. In brief, this intervention provided loans as well as other financial and non-financial services to rural poor households to invest into a cow or a buffalo. The aim of the intervention is to enable milk-selling as an extra income generating activity to households that typically depend on agriculture as their major source of income in an extremely drought-prone environment. In January/February 2008, 1,041 households living in 64 villages were interviewed. Of the respondents, approximately half were clients of this intervention. The other half of the sample was equally split among non-clients residing in the program villages and potential clients in non-program villages. Of the sample, 951 (91 per cent) of households were re-interviewed in the period April-June 2009.[5] The interested reader is referred to Augsburg (2009) for more details on the survey design and evaluation results from the first survey round.

Table 1 shows some summary statistics of the respondent households. The average household is headed by a married male, 45 years of age who has not received any formal education. In fact, only 27 percent of the household heads in our sample have more than primary education. The percentage for their spouses (who are on average four years younger than the household head) is even a bit lower.

The average household has five members, about two of them female and at least one younger than 16 years of age. About half of the households belong to the backward caste, 28 percent to the forward caste and 13 percent to the scheduled caste.

The primary activity of 63 percent of the households is agricultural labor and 25% are farmers, implying that at least 88% of all the respondent households are dependent on income from agriculture.

The survey included, in both waves, a number of questions aimed at eliciting some of the quantiles of the distribution of future household income. These questions on expected future income complemented a set of standard questions on current (actual) income.[6]

We followed a by now well established tradition in the phrasing of expectations questions, which starts by asking the respondents to report the range of variation of future income.[7] Once the range of variation has been established, the interviewer divides the resulting interval in four equal intervals by identifying cut off points A, B and C and then asks

[5] See Appendix A for a discussion of the issue that the second survey round was not collected exactly one year after the first.

[6] In what follows we will be referring to 'actual' or 'current' income interchangeably as the income earned by the household in the year previous to the interview (either 2008 or 2009). We will be referring to future income as the household income that will be earned by the household in the year following the interview. Therefore, in 2008, 'future' income is realized (and observed) in 2009.

[7] It is not obvious whether the reported values truly reflect the maximum (minimum) or some very high (low) quantile, see the discussion in McKenzie et al. (2011). In what follows, we treat these values as reflecting actual minimum and maximum income.

Table 1: Characteristics of the respondents households - 2008

Variable		2008 Sample		
		mean	med	s.d.
Household head	Age	44.5	45.0	11.8
	Gender	0.92	0.92	0.27
	No formal education	0.63	1.00	0.48
	Some primary education	0.10	0.00	0.31
Household composition	No of household members	4.73	4.00	1.72
	No of females	2.29	2.00	1.19
	No of kids age 0-5 yrs	0.35	0.00	0.64
	No of kids 6-10 yrs	0.47	0.00	0.782
	No of kids 11-16 yrs	0.65	0.00	0.81
	No of elderly (>63 yrs)	0.16	0.00	0.42
Caste of household	scheduled caste	0.13	0.00	0.34
	Scheuled tribe	0.05	0.00	0.21
	Backward	0.49	0.00	0.50
	Forward	0.28	0.00	0.45
Primary activity of household	Farmer	0.25	0.00	0.43
	Self-employed	0.06	0.00	0.24
	Agricultural labour	0.64	1.00	0.48

This Table provides descriptive statistics for the repondend and the corresponding household at the time of the first survey round, in 2008. Education, caste and primar6y activity variables are expressed as fractions.

for the likelihood of the respondent's income being higher than the threshold in the coming year. The exact wording of the questions is given in Appendix B. This question format has been used in other studies. Dominitz & Manski (1997b) use data where respondents were asked about four such thresholds. Nevertheless asking about more than one threshold is rare in developing countries, despite its benefits[8]. As the majority of respondents had no or very little schooling it could not be expected for them to know the concept of probabilities and the probability laws that these follow. Appendix A described how these concepts and rules were explained to the respondent.

As can be seen in Table 2, out of the 1,041 households to whom the questionnaire was administered in 2008, 1,012 (97 per cent) gave answers to the questions on minimum and maximum expected overall household income as well as to the questions on probabilities.

[8]Delavande et al (2011b) conduct a study in India where they compare different methods of eliciting subjective expectations.They find that precision improves if larger number of intervals are used.

Table 2: Response rates

		Round 1: 2008	Round 2: 2009
		(1)	(2)
Total no of observations		**1,041**	**951**
Information on Income provided		1,039	950
Min/Max provided		1,030	950
At least one probability missing		29	10
Wrong	Violation of monotonicity	5	17
	Wrong 'direction'	2	4
TOTAL no of observations available		**1005** (96.5%)	**919** (96.6%)

This Table provides information on responses to questions on income and income expectations in 2008 (column (1)) and 2009 (column (2)).

Of the 29 households whose probabilities were not elicited, 15 did give answers on the expected minimum and maximum income.[9]

A very similar pattern holds for the second survey round. The response error is on average slightly less but problems with readability of survey formats reduces somewhat the number of available observations, namely by four observations.

Overall, we have responses to income expectation questions from more than 96 per cent of the sample (the percentage is even higher for realizations) – these are rates that do not corroborate the common finding of substantial non-response for income questions.

The respondents' willingness to respond does not necessarily imply meaningful answers though. Before using the subjective expectation data in statistical model for income, it remains important to validate responses and to judge whether expectations are reported coherently.

Given the environment the respondents live in and the uncertainty they face one would

[9]We analyze item non-response in unreported regressions. The respondents that did not give any answers do seem to own significantly less land but are not significantly different in key characteristics such as education level, caste, primary activity of household, household composition and wealth (savings and assets) of the household. The only other significant variable is the number of male household members, which has a positive coefficient.

think that the concept of probability and risk should be salient to them. Nevertheless, as elaborated previously, an important concern in the elicitation of probabilities is that respondents not trained in probability theory, may find it difficult to answer the specific questions we pose and formalize their subjective perceptions about uncertain events (see for example Walley (1991)). In an earlier working paper version (2012), we validate in detail whether the answers that were provided to the set of expectation questions make sense and argue that the answers provided do reflect respondents' believes. We provide here a brief summary:

1. **Logical Response Errors**: Numbers in the bottom panel of Table 2 show that in both survey rounds, violations to basic probability laws (monotonicity, and wrong 'direction') make up less than one per cent of the sample.[10]

2. **Bunching of Percentages**: Table 3 reports the answers given when asked for the percent chance that next year's income will be less than the different thresholds.[11] We can see that only a negligible number of respondents give 0 or 100% as answers, which holds for both survey rounds. This gives confidence in the elicited range. A good half of the sample reports a 50% chance for the midpoint, an issue we will take up again later. We further note that while respondents had a tendency to round to the nearest 5% (hardly any respondents use number such as 23%, 71%) and that the whole range used. One explanation for the rounding to the nearest 5% is the design of the visual aid (the ruler) which was used to elicit probabilities. The ruler had marks only for steps of ten and these numbers were written on the ruler. Respondents might

[10]This percentage is much lower than what is found in other studies. Dominitz and Manski (1997a), for example, find that almost five per cent of their respondents violate – the rate is more than double when one includes respondents who initially gave an answer violating monotonicity and were prompted for a revision. Such prompting was not allowed in our study. However, interviewers were asked to prompt during the rainfall questions. A considerable amount of time was spent on explaining the numbers zero to 100 represent and on practicing using such a scale with the rainfall questions. Unfortunately it was not recorded how quickly respondents picked up the idea of probabilities or whether they made mistakes during the rainfall questions. We can therefore not correlate these potential indicators of how well the probability concept was understood with the answers provided for the income expectations.

[11]If many respondents answer 0% for the lowest and 100% for the highest threshold, then this is indication that either the elicited range of possible future income realizations is wrong or that the concept of the questions was not understood in general – both undesirable for the use of the expectation data. Finding many respondents to answer 50% for the middle threshold is not a problem per se but has implications for the distributional assumptions one has to make later in the construction of the subjective income distribution.

therefore have been induced to point to these marks instead of somewhere in-between them.

Table 3: Reported probabilities 2008 and 2009

	2008			2009		
	0%	50%	100%	0%	50%	100%
Threshold A (lowest)	0	2	7	0	21	14
Threshold B (midpoint)	0	544	0	0	147	0
Threshold C (highest)	1	51	0	1	108	0

This Table shows the number of resopndents which reported 0%, 50%, and 100% in both survey rounds.

Taking the negligible percentage of logical response error, the sensible pattern of probabilities for different thresholds, and the expected correlations between stated probabilities, one can be relatively certain that the answers provided conform to the basic laws of probability and that respondents did not give some random answers for the sake of answering.

We now turn to construct an individual specific subjective income distribution whose moments we can then relate to realized income and other households characteristic to give further evidence for the reasonableness of this direct subjective information.

3 Fitting a Subjective Income Distribution

If one believes the conclusion from the validation exercise above, then one can interpret answers to the percentage chance questions as points on the subjective cumulative distribution function of future household income. With that, one can fit a respondent-specific subjective income distribution, which can subsequently be used to compute income moments and to analyze how income expectations vary with respondents' realized income. In this section, we first make an assumption about the distribution of future income. We then proceed to compute some moments and quantiles that that particular distribution implies and compare them to statistics of actual income.

3.1 Piece-wise uniform distribution

To fit the income distribution, we assumes a piece-wise uniform probability distribution and focus on the means and standard deviations of these distributions. One could use other, more complicated distribution functions. However, the three points of the c.d.f. elicited from each respondent are not sufficient to determine which distribution fits best the shape of the respondent's subjective outcome distribution. Dominitz and Manski (1997a) – with information on four points on the cdf - assume a log-normal income distribution fitted via non-linear least square. The assumption of log-normality is a very common one in studies of realized income. McKenzie et al. (2013), for instance, find that a log-normal distribution fits well the distribution of income in their Tongan sample. Nevertheless, a log-normal would be an inconsistent choice with data that point to a right-skewed income distribution. As will be seen below, for many households, the data do display skewness to the right for the first survey round (the expectation data we will need to use to compare realizations and expectations over the same period). Furthermore, as already alluded to in the analysis of bunching of percentages, many households stated 50% for the probability of earning at least the midpoint of their expected range, which is not in line with the assumption of log-normality. Because of these two observations, we decided to assume a piece-wise uniform which can, if needed, be interpreted as an approximation to more complex distributions.

Given that in the final section of the paper we use mainly expectations in the first survey round, we concentrate our discussion here on the fitted subjective income distribution for 2008.[12] In particular, Table 4 shows the average probabilities assigned by respondents in 2008 to the four sections in which the range of possible values for future income is divided (Min to A, A to B, B to C and C to Max - where we recall that Min and Max are elicited from the respondents and B=(MIN+MAX)/2, A=(MIN+B)/2 and C=(B+MAX)/2).

The first section (MIN to A) has the lowest mean (and median) probability and the

[12]Results for 2009 are displayed in Table 13 and Figure 4 in Appendix C.

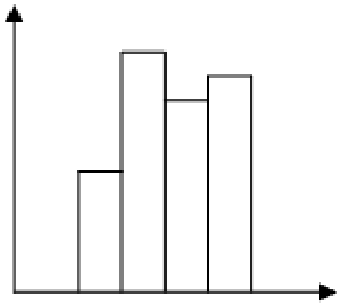
	obs.	min	max	median	mean	std.dev.
Min - A	1005	0	50	10	16.13	8.48
A - B (Midpoint)	1005	10	60	30	30.81	8.28
B (Midpoint) - C	1005	0	57	30	26.11	6.87
C - Max	1005	0	70	30	26.95	9.71

Table 4: Probabilities assigned to sections of income distribution, 2008

This Table displays descriptives statistics for the probabilities assigned to the four equally spaced intervals of the income distribution from the first survey round in 2008.

second section (A to B) the highest. Income appears slightly right skewed, while the probabilities of the two intervals among the midpoint add up to the highest value (around 57 per cent). Figure 1 gives an idea of the average individual income distribution derived from the answers to the expectations questions. The results for the second survey round, reported in the Appendix, differ from those in Table 4 in that the distribution has its highest probability mass in the fourth interval, which leads to a left-skewness of the data.

Figure 1: The Piecewise Uniform Distribution, 2008



3.2 Moments of the distributions of future income: uncertainty and heterogeneity.

Having fitted a probability distribution for each individual in the sample, we now look at different moments and quantiles of the subjective distribution of future (log) income. We display the summary statistics for these moments in Table 5. (Appendix C, Table 14 gives provides the same information for income variables in levels.)

We discuss at length the relationship between actual and expected income in the next section. In that section we will also discuss the variable'Normal income' and how this was elicited. Here we only note that their cross sectional means are very similar, both in 2008 and in 2009. The cross sectional standard deviation of actual (current) and expected income is also comparable, although the one of actual income (at 0.780) is a bit higher than that of expected income (at 0.738). This difference is consistent with the fact that the variability in the cross section of actual income reflects the influence of unanticipated shocks that, obviously are not reflected in expected income.

These considerations lead us to the analysis of higher moments of the distribution of future income and, in particular, the analysis of uncertainty. An interesting aspect to note in Table 5 is that we can now distinguish between heterogeneity and uncertainty. Much of the literature, which does not have access to subjective expectations data, is forced to use the cross sectional variability as a proxy for uncertainty. Obviously this is legitimate only under very stringent assumptions. A similar point is made, in the context of a consumption insurance model with permanent and transitory components of income by Kaufmann and Pistaferri (2009) who consider both 'anticipated' and 'unanticipated' changes in income.

As we mentioned in the introduction, without the data on subjective expectations we would be forced to make inferences about the size of uncertainty from the realizations of income in the cross section, which are affected by stochastic elements but also by heterogeneity that might be unobserved to the econometrician but not constitute an element of uncertainty for the individuals in the sample. In 2008, for example, we find the cross sectional mean of the the standard deviation of log income computed from the subjective expectations variables is 0.164. The standard deviation of actual log income in the cross section is instead 0.780, a much larger number which reflects both uncertainty and heterogeneity.[13]

[13]Of course one could model some of this heterogeneity as driven by observables. Here we are simply making the point in a stark fashion.

Table 5: Summary statistics of income variables (2008 and 2009, in logs)

	2008 - Round 1					
Variable	Obs	min	max	p50	mean	sd
Expected Income:						
Mean	988	7.485	13.627	10.924	10.921	0.738
Standard deviation	990	0.006	1.017	0.144	0.164	0.127
Coeff. of Variation	988	0.000	0.103	0.013	0.015	0.012
Min	1,021	7.601	13.122	10.597	10.600	0.793
Max	1,022	8.006	13.592	11.156	11.155	0.709
Range(Max-Min)	1,021	0.021	2.878	0.511	0.558	0.326
Realized Income:						
Current income	1,033	8.825	13.400	10.897	10.910	0.725
Normal income	1,024	7.601	13.305	10.820	10.802	0.780
	2009 - Round 2					
Variable	Obs	min	max	p50	mean	sd
Expected Income:						
Mean	906	8.396	13.199	10.969	10.903	0.659
Standard deviation	906	0.004	0.471	0.152	0.151	0.062
Coeff. of Variation	904	0.000	0.043	0.014	0.014	0.006
Min	944	8.006	13.816	10.597	10.552	0.714
Max	947	8.294	13.199	11.225	11.110	0.666
Range (Max-Min)	935	8.517	13.199	11.225	11.118	0.648
Realized Income:						
Current income	944	8.476	13.286	11.007	10.981	0.605
Normal income	943	8.476	12.948	11.019	10.983	0.603

This Table shows descriptive statistics of the natural logarithm of expected and realized income variables elicited in both survey rounds. The variable "Normal income" is described in Section 4.3.

3.3 Comparison of expected and actual income

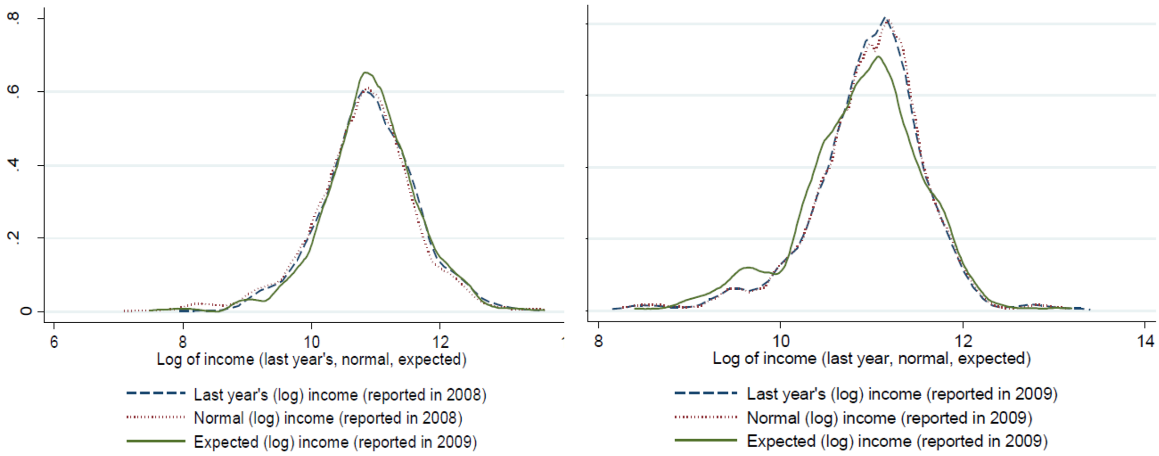
Having described the main features of the moments of the subjective income distribution we now focus on the mean of the distribution and look at how expected (log) income covaries with actual income.[14] A similar exercise is performed for US earning data, by Dominitz (1998).

Figure 2 gives a first glance of the extent to which expectations and realizations move together in the two survey rounds. In particular, for each survey year, we plot the cross sectional distribution of actual and expected income. In addition, we also plot what we define as 'normal' income, which is a question asked to the respondents after asking their

[14] The analysis of higher moments is currently under way and will be discussed in a separate paper.

current income.[15]

Figure 2: Normal, last year's and expected household income, 2008 and 2009



The main conclusion to be drawn from these graphs is that the distribution of expected and realized incomes (and normal income) seem very similar. Moreover, expected and realized income measures are strongly correlated in the cross section. In Table 6, we report the correlation between expected income (as computed from the elicited subjective expectations and the assumption made on the probability distribution) and income realizations (current as well as normal income measures). The correlations in both years are positive (between 0.764 and 0.873) and highly significant (significance levels are displayed in brackets), but statistically different from each other. Finding very different distributions of realized and expected income would not have been proof of expectation data being wrong, but finding them moving closely together can be seen as support for their salience and validity.

We have already seen that the distribution of expected future income and actual income realization are reasonably similar, and that they co-vary together in the cross section. As

[15]The way we collected this information is as follows: households were asked about the amount they earn for all their different income sources. This information was summed up by the interviewer and read out to the respondent. The respondent was then asked whether this total income from the previous year is a typical income and if not, what a typical income would be.

Table 6: Correlations of expectations with realizations

		2008	2009
Expected income &	Normal income	0.764 (0.000)	0.873 (0.000)
	Last year's income	0.848 (0.000)	0.861 (0.000)

This Table shows correlations between average expected income and realized income measures (normal and last year's) in the cross section. Significance levels of the correlations are displayed in brackets.

an additional validation exercise of our expectations data, we check whether they co-vary with observable variables in a similar fashion. Table 7 shows the estimates of the regression coefficients (standard errors clustered at village level and bootstrapped) we obtain relating, in turn, current and expected log income to covariates. We note that since we do not attempt for any possible endogeneity of explanatory variables, the presented information does not say anything about income determination but serves a descriptive purpose.[16]

The first thing to observe from Table 7 is that estimated coefficients are mostly in line with what one would expect. If we consider, for example, those variables that are significant at the conventional level of 5 percent we see that if the household head has no formal education, then income is reduced significantly and these households with a non-educated household head also expect to have a lower income in the future. Households headed by a married individual have significantly higher income (both realized and expected).Households that belong to a lower caste or minority group also earn lower income, and expect to do so, and farmers earn significantly more than households with other primary sources of income. The same holds for households with more female household members.

It is interesting to note that some variables are significant only in one of the two regressions presented. Most notable is the estimated coefficient for whether the household has children in the age range 11 to 16 years. While this does not correlate significantly with realized income, having children of soon-to-be working age in the household seems to have

[16]We also run the regression with the midpoint between reported minimum and maximum expected income as the dependent variable, which can be seen as a 'raw' measure of expectations, not relying on any functional form assumption. Results differ only marginally from those reported and where we make distributional assumptions and are available upon request.

Table 7: Realized and expected income on covariates

		Dependent variable (logs) Current Income (1)	Average E[Income] (2)	p-value diff of coeff (3)
Characteristics of the household (hh) head	Age of hh head	0.017* (0.010)	0.005 (0.012)	(0.041)
	Age^2 of hh head/100	-0.015 (0.011)	-0.001 (0.013)	(0.028)
	Hh head is male (0/1)	-0.052 (0.087)	-0.043 (0.126)	(0.877)
	Hh head no education (0/1)	-0.157*** (0.036) (0.041)	-0.183***	(0.256)
	Hh head is married (0/1)	0.353*** (0.094)	0.353*** (0.118)	(0.996)
Household Composition	No of female hh members	0.064*** (0.020)	0.066*** (0.019)	(0.884)
	No of kids age 0-5	-0.022 (0.031)	-0.045 (0.031)	(0.187)
	No of kids age 6-10	0.021 (0.022)	0.030 (0.023)	(0.559)
	No of kids age 11-16	0.035 (0.021)	0.047** (0.023)	(0.330)
	No of elders	-0.017 (0.048)	-0.049 (0.046)	(0.134)
Caste of the Household	Backward	-0.085* (0.048)	-0.034 (0.057)	(0.024)
	Scheduled caste	-0.159* (0.081)	-0.167* (0.086)	(0.857)
	Scheduled tribe	-0.295*** (0.076)	-0.240*** (0.101)	(0.220)
	Minority	-0.246** (0.107)	-0.108 (0.106)	(0.019)
Primary Activity of hh	Farmer	0.326*** (0.060)	0.408*** (0.078)	(0.086)
	Agricultural labourer	0.057 (0.050)	0.105* (0.062)	(0.290)
Household lived in village all their life		0.018 (0.081)	0.013 (0.083)	(0.948)
Constant		10.074*** (0.294)	10.241*** (0.357)	(0.)
1-5 Sample Size		1,829	1,751	
Adjusted R^2		0.091	0.097	

This Table displays the estimated coefficients of regressing income measures on covariates. Column (1) gives results where the dependent variable is current income and in column (2) the dependent variable is calculated average expected income. All standard errors (in brackets) are clustered at the village level and bootstrapped. Stars indicate significant coefficients at the conventional 10, 5 an 1% significant levels.

a significant and positive relationship with expected income.

We present in the last column of the Table the p-value for the test of equality of the coefficients in the two specifications. It can be seen that only for the age of the household head and for indicators of whether the household belongs to the backward caste or minority do the coefficients differ significantly across the two regressions. Belonging to a backward caste or minority for example has a significantly larger coefficient on realized income, whereas the relationship is much smaller when considering the expected income as reported by these households.

It is reassuring to observe that both realizations and expectations of income vary sensibly with covariates. By and large, the patterns of correlations are very similar, although for some variables, the coefficients in the regressions for expected income and actual income are significantly different in size.

Having considered that expectations and realizations vary sensibly with covariates, we turn to modeling the income processes.

4 Modeling income processes

In this section, we propose a simple statistical model for income. There is a voluminous literature on modeling the stochastic processes for income, earnings or wages in developed countries, going back to Friedman (1954), Friedman and Kuznet (1954) and Lillard and Willis (1978). More recent and often cited contributions include MaCurdy (1982, 1983), Abowd and Card (1989), Gottshalk and Moffitt (1997), Meghir and Pistaferri (2004), Alvarez and Arellano (2003), Guvenen (2007). The evidence on developing countries is much scantier.

All the papers mentioned above use longitudinal data to estimate the stochastic properties of income. In what follows, instead, we make use of the subjective probabilities data to estimate such a process.[17] If expectations are rational, it should not make any difference whether one uses actual income realizations over time or income realization and expectations data. The latter approach, however, changes the nature of the residuals and, consequently, the nature of the econometric techniques one uses.

[17] A paper that performs a related exercise is Dominitz (2001), which estimates the relationship between income expectations and current income (as well as other variables). The main purpose of that paper is to study this relationship and a model of expectations formation rather than a model for income, as we do.

We start with a very simple model that we extend momentarily. Suppose that (log) income is given by the following equation:

$$y_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \beta x_{i,t} + u_{i,t} \tag{1}$$

where $y_{i,t}$ is realized (log) income for household i at time t, $x_{i,t}$ are household characteristics, such as the age of the household head, household composition and primary activity of the household, and $u_{i,t}$ is an $i.i.d.$ innovation to the income process. We assume that the variables $x_{i,t}$ evolve in a deterministic fashion. Under rational expectations, the expectations of $y_{i,t}$ conditional on information available at time $t-1$, is given by:

$$y^e_{i,t} = E[y_{i,t}|y_{i,t-1}] = \alpha_0 + \alpha_1 y_{i,t-1} + \beta x_{i,t} \tag{2}$$

Let's now denote with $y^{ee}_{i,t}$ the expected value of $y_{i,t}$ as computed from the subjective expectations data. We define the difference between $y^{ee}_{i,t}$ and $y^e_{i,t}$ as $v_{i,t}$. $v_{i,t}$ can effectively be interpreted as measurement error in the subjective expectations data or as a deviation from rational expectations. If we assume that this term is uncorrelated with realized current income or the background variables $x_{i,t}$, one can estimate the parameters of the income process using the following regression:

$$y^{ee}_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \beta x_{i,t} + v_{i,t} \tag{3}$$

Notice that now $v_{i,t}$ is not an innovation but reflects measurement error. Notice also that a simple model such as that in equation (1), where the income process is Markov, can be estimated with a single cross-section.

An important extension of the model in equation (1) is to allow for the possibility of fixed individual effects f_i:

$$y_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \beta x_{i,t} + f_i + u_{i,t} \tag{4}$$

Fixed effects are important from a statistical and economic point of view, as their presence changes substantially the interpretation of the persistence one observes in the data. Moreover they imply that a simple OLS estimate of equation (3) would yield inconsistent estimates of the coefficients of interest.

We propose different solutions to deal with the possibility of fixed effects. First, we use a survey question that asks respondents to report 'normal income' and assume that fixed effects are a function of such a variable. Therefore, instead of equation (3), one can estimate:

$$y^{ee}_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \beta x_{i,t} + \gamma y^N_i + v_{i,t} \tag{5}$$

where y^N_i is normal income. Notice that, in principle, such an equation can also be estimated with a cross-section only.

In what follows, we use the measure of 'normal income' we introduced earlier. As indicated in the notation, we assume that 'normal income' is fixed over time. An issue that we need to face, in this case, is that the answers the respondents provide to this question change from one survey to the next. We discuss how we deal with this issue below.

A second approach we take to deal with the presence of fixed effect is standard in the literature on dynamic panel data and on estimating income processes. In particular, we use the availability of a longitudinal dimension to estimate the income process in first differences. The fact that the fixed effects appear additively in our model implies that they difference out. From equation (3) we derive:

$$y^{ee}_{i,t} - y^{ee}_{i,t-1} = \alpha_1 (y_{i,t-1} - y_{i,t-2}) + \beta (x_{i,t} - x_{i,t-1}) + v_{i,t} - v_{i,t-1} \tag{6}$$

Under rational expectations, the error term $v_{i,t} - v_{i,t-1}$ reflects only measurement error in the expectations and we are able to estimate this equation with simple OLS.

As yet another alternative, we can adapt the technique due to Arellano and Bover (1995)[18]. In particular, we can consider the following regression:

$$y^{ee}_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \varepsilon_{i,t} \tag{7}$$

where $\varepsilon_{i,t} = v_{i,t} + f_i$ and where we omit the $x_{i,t}$ for notational simplicity. Under the assumption that the correlation of the fixed effect f_i with $y_{i,t-1}$ and $y_{i,t-2}$ is the same, one can use $y_{i,t-1} - y_{i,t-2}$ as an instrument for $y_{i,t-1}$ and obtain consistent estimates of the parameter of interest.

There are three final specification issues we need to discuss as they are relevant for all specifications we have considered so far. First, given the nature of the data, it is not unlikely that actual income is affected by measurement error. Income is notoriously difficult to collect in developing countries. Second, the literature on income processes we cited above, often removes in a first stage regression the effect of the 'deterministic' variables $x_{i,t}$ and studies the residuals from this regression. Third, as we mentioned above, the residuals $v_{i,t}$ in equation (3) represent either measurement error in expectations or deviations from rational expectations. As such, they might be correlated with observable variables $x_{i,t}$ and induce a bias in our estimates. For instance, less educated individuals might make systematic errors in answering the expectations questions.

As for the first issue, we use the fact that our data comes from 62 villages to use village level averages as instruments for actual income (or changes in income). As for the second issue, we follow the standard practice and remove the effect of $x_{i,t}$ variables. The results are barely affected and those where the $x_{i,t}$ variables are maintained in the income equation are available upon request. Finally, there is not much we can do about the third problem.

[18] We thank Manuel Arellano for this suggestion

However, we point out that in our first difference specifications, all observable variables that are constant over times (such as education) drop out of the equation. Moreover, to cause a bias, the measurement error in the expectations variables should be related to the observable variables in the *mean.*

5 Estimating income processes.

In this section, we report the results we obtain estimating the specifications discussed in Section 5 on our data. We first present results for specifications in levels and first differences. In all cases, we will report both estimates obtained by OLS and estimates obtained by IV to deal with measurement error. We then move on to discuss results obtained when applying the Arellano and Bover (1995) technique. Before doing so, we need to discuss in a bit more detail our measure of normal income, which will be used in all specifications as one way of dealing with the presence of fixed effects.

5.1 *Normal* Income

As mentioned in the previous section, we assume that normal income is fixed over time so that we need to deal with the issue that the answers provided by respondents to the questions relating to their normal income vary from one survey to the next.

This can be seen in Table 8. Changes in normal income have a mean of Rs. 2,735 with a standard deviation of 65,365. The median difference lies at Rs.8,400 (Summary statistics of the all income variables in levels are displayed in Appendix C in Table 14).

We can see that the assumption of normal income being constant over time is not fully met. This is not too surprising as it is likely that household normal income depends on some time-varying variables such as household composition. In Table 9 we report the

Table 8: Normal income

	obs.	min	max	median	mean	std.dev.
Difference: Typical income 2009-2008	931	-544,000	348,400	8,400	2,735	65,365
As above (ln of absolute value)	915	5.298	13.487	10.342	10.222	1.084
Growth of typical income from 2008-2009	931	19.00	0.960	0.156	-0.14317	1.828

This Table displays descriptive statistics of three vearibales constructed based on the 'normal income' reported by the households.

results obtained estimating a regression of growth in normal income between 2008 and 2009 on four indicator variables: whether the household has more kids in the age range 0-5, 6-10, 11-16 in 2009 than in 2008 and whether there are more elders (household members older than 63 years) or not. As in all other regressions, standard errors are bootstrapped and clustered at the village level.

While the explanatory power is low, we can see that an increase in number of children in the age ranges 6 to 10 and 11 to 16 from 2008 to 2009 results in a significant increase in the growth of normal income over the same time period. On the other hand, while insignificant, an increase in the number of small children (in the age range 0-5) reduces the growth of normal income. These results are in line with the hypothesis that an additional household member implies additional labor and hence higher total household income – unless the additional member is too young to work (and might take away labour hours from, for example, the mother).

Given this dependence on some time-varying characteristics, we decide to take the average of typical income over the two years of data we have and use this as a measure of normal income and hence as a proxy for the fixed effect likely to be present in equation (1).

5.2 Level and first difference specifications

The first three panels (titles "2008", "2009" and "Pooled") of Table 10 report estimates of the coefficients α_0 and α_1 in the model in equation (3), which takes log expected income as the dependent variable. As mentioned above, our estimates were obtained from a specification where the effect of the $x_{i,t}$ variables was eliminated in a preliminary regression.

Table 9: Normal income on covariates

Regress: Growth in typical income (2008-2009)	Coeff (std.err.)	Coeff (std.err.)	Coeff (std.err.)
Indicator for...			
...more kids age 0-5 in 2009 than in 2008	-0.141 (0.201)	-0.0118 (0.189)	-0.247 (0.185)
...more kids age 6-10 in 2009 than in 2008	0.274** (0.115)	0.258** (0.121)	0.236* (0.123)
...more kids age 11-16 in 2009 than in 2008	0.235* (0.125)	0.210* (0.123)	0.123 (0.127)
...more elders (>64yrs) in 2009 than in 2008	0.073 (0.149)	0.107 (0.155)	0.171 (0.157)
...became landless between rounds		-0.306 (0.193)	-0.366** (0.181)
...became landowner between rounds		0.429*** (0.111)	0.333** (0.132)
...lost livestock between rounds			-0.604*** (0.191)
...became livestock owner between rounds			0.353*** (0.124)
...lost irrigation equipment between rounds			-1.027*** (0.0.297)
...became irrigation equipment between rounds			0.003 (0.220)
Constant	-0.414*** (0.112)	-0.453*** (0.106)	-0.171 (0.206)
R^2	0.007	0.018	0.073

This Table shows results from regressing the growth in normal income in the two survey rounds on indicators for the households' composition. Standard errors are clustered at the village level.

This procedure, however, does not affect much the results we obtain.

The coefficient α_1, that captures the persistence of the income process, is estimated at 0.872 with the 2008 survey data with a standard error of 0.031 (column (1a)). We can therefore reject the hypothesis that the coefficient is equal to 1, which would indicate a random walk. The same conclusion is drawn for the data collected in 2009, as well as when pooling the two years together, regardless of whether we control for normal income in the specification (column (2a)) or not (column (1a)).

However, if we instrument realized income using average income in a village, the esti-

Table 10: Income process: Level specification - OLS and IV

Dependent variable: Expected (ln) income		OLS (1a)	IV (1b)	OLS (2a)	IV (2b)
2008	Income last year (ln)	0.872 (0.031)	0.955 (0.042)	0.832 (0.046)	0.967 (0.076)
	Typical income (ln) (AVG)			0.115 (0.056)	-0.003 (0.081)
	Constant	-0.004 (0.012)	-0.005 (0.012)	-0.005 (0.015)	-0.005 (0.014)
	F-stat. (1st stage)		5437		153
	Prob>F		0.000		0.000
2009	Income last year (ln)	0.844 (0.049)	1.280 (0.083)	0.832 (0.047)	1.544 (0.177)
	Typical income (ln) (AVG)			0.065 (0.049)	-0.469 (0.149)
	Constant	0.006 (0.025)	0.008 (0.018)	0.004 (0.026)	0.003 (0.022)
	F-stat. (1st stage)		2938		99.93
	Prob>F		0.000		0.000
Pooled Sample	Income last year (ln)	1.030 (0.007)	1.009 (0.022)	1.079 (0.020)	0.882 (0.291)
	Typical income (ln) (AVG)			-0.069 (0.028)	-0.167 (0.352)
	Constant	-0.341 (0.072)	-0.120 (0.244)	-0.127 (0.121)	-9.595 (0.691)
	F-stat. (1st stage)		7.7e+12		13.52
	Prob>F		0.000		0.081
Differenced	Income (ln) 2009-2008	0.862 (0.032)	1.056 (0.056)	0.682 (0.034)	1.044 (0.058)
	Typical income (ln) (AVG)			-0.057 (0.069)	-0.033 (0.076)
	Constant	0.010 (0.244)	0.010 (0.020)	0.008 (0.023)	0.008 (0.020)
	F-stat. (1st stage)		1268		1179
	Prob>F		0.000		0.000

This Table shows results from estimating equation(3) in the first three panels (titled "2008", "2009" and "Pooled"). The lower panel (titled "Differenced") shows results from estimating equation(**??**). Columns (1a) and (2a) show simple OLS regressions, where in (1b) and (2b) realized income is instrumented with average income in a village. The F-statistics shown in these columns are from the first stage regression, testing significance of the instrument. The effect of the $x_{i,t}$ variables was eliminated in a preliminary regression. Standard errors ares clustered at the village level.

mates of the coefficient on current income increases (as shown in columns (1b) and (2b)) and, as a consequence, we cannot reject the hypothesis of a random walk. Using data from 2009 we marginally reject the hypothesis that α_1 is equal to unity. However, using the 2008 data and both years pooled, we are not able to reject the hypothesis that the coefficient α_1 is equal to unity anymore. In 2008 α_1 =0.955 with a standard error of 0.042 and pooling the data α_1 =1.009 with a standard error of 0.002. This implies that income follows a random walk.

As we have discussed above, since we have two years of data available, we can account for the presence of fixed effects in another way. By taking the first difference of our expectation equation we difference-out the fixed effects, as reflected in (**??**). These results are shown in the lower panel of Table 10, titled “Differenced”.

The results are in line with those just presented: Without instrumenting the difference in realized income (columns (1a) and (2a)), we reject the hypothesis that α_1 is equal to one, whereas we are not able to do so when using aggregate village income information as an instrument (columns (1b) and (2b)). The coefficient on α_1 is in that case estimated to be 1.056 with a standard error of 0.056 without including normal income in the estimation and α_1 =1.044 with a standard error of 0.058 when including this information.

5.3 Arellano-Bover method.

The previous section discusses another possibility to estimate the parameters of the income process, namely by using the approach proposed by Arellano and Bover (1995) and instrument $y_{i,t-1}$ in equation (7) with $y_{i,t-1} - y_{i,t-2}$. We report the results we obtain using this approach in Table 11. In column (1) we obtain a coefficient of 0.833 which is marginally significantly different from 1. In column (2) we add to the specification ‘normal’ income, whose coefficient turns out to be not statistically different from zero. The coefficient on current income is virtually unchanged.

The evidence we have presented indicates that the income process faced by the households in our sample is extremely persistent. In many specifications we cannot reject the hypothesis of a random walk. We also find that our IV specification yield slightly larger

Table 11: Income process: Arellano-Bover estimator

Dependent variable: 2010 Expected (ln) income	Coeff. (1)	(2)
2009 Income (ln)	0.833	0.835
	(0.054)	(0.056)
Typical income (ln) (AVG)		0.063
		(0.048)
Constant	0.003	0.004
	(0.026)	(0.027)
No. obs.	843	832
F-stat. (1st stage)	188	532
Prob>F	0.000	0.000

This Table shows results from estimating equation(7), instrumenting realized income with average income in a village. The effect of the $x_{i,t}$ variables was eliminated in a preliminary regression. Standard errors ares clustered at the village level. The F-statistic shown is from the first stage regression, testing significance of the instrument.

point estimates of the persistence parameter, indicating the presence of attenuation bias, probably induced by measurement error in current income. These findings are robust across specifications. In particular, we obtain them both in levels and first difference specification. The robustness of the result also assuage the worry that the results we obtain are driven by a correlation between measurement error in the expectation variables and other observable variables. A bias of this kind would yield different estimates when moving, for instance, from level to first difference specifications.

6 Conclusions

In this paper, we have analyzed data on subjective expectations about future household income elicited in a survey conducted in 2008 and in 2009 in the district of Anantapur, India. Although the survey respondents are very poor and with little formal schooling, we show that the answers to the expectations questions are, by and large, internally consistent and sensible. The first contribution of the paper, therefore, is the validation of the subjective expectations data. We find that not only respondents were willing to answer the questions but they answered in a way not inconsistent with the laws of probability theory.

Given the answers to the subjective expectation questions and a simple assumption about the probability distribution of future income we can compute expected future income and use it, along with information on current income, to estimate a time series model of the income process. Having computed expected future income, partly to further validate our data, we compare it to actual income and relate it to a number of observable variables. We find that expected income varies with observables in a way similar to actual income. This evidence confirms our impression that the data are of good quality and measure actual income expectations. The structure of the data additionally allows us to infer the size of income risk individual households face and compare it to the (cross section) variability of income in the sample. Notice that the variance of income computed from our expectations questions should reflect individual uncertainty, while the variability of observed income across individual households in the cross section will reflect uncertainty, predictable (by the agent) changes and heterogeneity. It is important to stress that some version of the latter is what is often used as a proxy for uncertainty in the absence of expectations data. We find that, as is to be expected, the cross sectional variance of (log) income is much larger than the variance computed from expectations data. In 2008, the mean of the second moment of the constructed standard deviation is 0.164. This compare to a standard deviation of actual income in the cross section of 0.780. (Notice that this latter number is not very different from the standard deviation of the mean expected (log) income, which we estimate at 0.738.) In 2009 we find similar results: the mean of the computed standard deviation (in logs) is 0.151 and the standard deviation of the actual income is 0.603.

In modeling income, we have adopted a framework that has been often used in the study of earnings dynamics, which relates current income to past income and shocks. Under rational expectations, this model would imply that future expected income depends on current income. We also allow for persistence induced by individual fixed effects. To deal with them, we use three different approaches: (i) we use a question on 'normal' income;

(ii) we use both waves of the survey we have to difference out the fixed effects and (iii) we use an instrumental variable approach proposed by Arellano and Bover (1995). Finally, we also allow for measurement error in current income, using an IV approach, where we use village averages to instrument individual income.

The results we obtain are remarkably consistent across specifications and indicate that income is extremely persistent. When estimated by OLS, that is ignoring the attenuation bias induced by measurement error, we obtain a coefficient just above 0.8. When we instrument current income with village level averages, instead, we obtain estimates very close to (and not significantly different from) one. This pattern is consistent with the presence of measurement error in current measured income that induces attenuation bias in the OLS estimates of the persistence coefficient.

To the best of our knowledge, our exercise is the first in which data on subjective expectations are used to estimate a model of income dynamics. Some of our approaches circumvents the necessity of longitudinal data. Our exercise shows that data on expectations can be collected even in the context of developing economies and can be used to estimate time series model for income that would not be identifiable in the absence of panel data.

7 References

Abowd, J., D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica* 57, 411-45.

Alessie, R., Lusardi, A., (1997). Saving and income smoothing: evidence from panel data. *European Economic Review* 41, 1251–1279.

Alvarez, J. and M. Arellano, (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica*. Econometric Society, vol. 71(4), pages 1121-1159.

Antman, F. and D. McKenzie (2007). Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity, *The Journal of Development Studies*, Taylor and Francis Journals, vol. 43(6), pages 1057-1083.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics*, vol. 68(1), pages 29-51,

Attanasio, O. (2009). Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review*. Apr 2009, Vol. 99, No. 2: Pages 87-92 .

Attanasio, O. and V. DiMaro (in progress). Income Expectations in Rural Mexico. Unpublished working paper.

Attanasio, O., C. Meghir, and M. Vera-Hernández (2005). Elicitation, validation, and use of probability distributions of future income in developing countries. Unpublished working paper, University College London.

Augsburg, B. (2009), Microfinance - Greater Good or Lesser Evil?, PhD dissertation, Maastricht University.

Das, M. and A. van Soest, 1997, Expected and Realized Income Changes: Evidence from the Dutch Socio-Economic Panel, *Journal of Economic Behavior and Organization*, 32, pp. 137-154.

Das, M., A. van Soest (1999). A panel data model for subjective information on houseold income growth. *Journal of Economic Behavior and Organization*, Vol. 40 (1990) 409-426.

Delavande, A., X. Gine, and D. McKenzie. (2011). Eliciting Probabilistic Expectations in Developing Countries: How Sensitive are Answers to Variations in Elicitation Design?. *Journal of Applied Econometrics*, 2011, 26, 479-497.

Dominitz, J. (1998), Earnings Expectations, Revisions, and Realizations, *The Review of Economics and Statistics*, Vol 80(3), pages 374-388.

Dominitz, J. (2001), Estimation of income expectations models using expectations and realization data., *Journal of Econometrics*, Vol 102, pages 165-195.

Dominitz, J., Manski, C.F. (1996), Eliciting Student Expectations of the Returns to Schooling. *Journal of Human Resources* 31:1–26.
Dominitz, J., Manski, C.F. (1997a), Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92(439), 855-67.
Dominitz, J., Manski, C.F. (1997b), Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations, *Public Opinion Quarterly*, 1997b, 61, 261-287.
Friedman, M. (1957). *A Theory of the Consumption Function*, Princeton University Press.
Friedman, M., Kuznets, S. (1954). Incomes from Independent Professional Practice, National Bureau of Economic Research, New York (1945), 1954.
Gottschalk, P., Moffitt, R. (1997). Can Families Smooth Variable Earnings. *Brookings Papers on Economic Activity* 1: 229-303.
Guvenen, F. (2007). Learning Your Earning: Are Labor Income Shocks Really Very Persistent?. *American Economic Review*, 97(3), 687–712.
Jalan, K. and M. Ravallion. (2004). Household Income Dynamics in Rural China, in Stefan Dercon (ed) *Insurance Against Poverty*, Oxford University Press.
Kaufmann, K. and L. Pistaferri (2009), Disentangling Insurance and Information in Intertemporal Consumption Choices, *American Economic review*, Papers & Proceedings, Vol. 99(2) 387-392.
Lillard, L. A., Willis, R.J. (1978). Dynamic Aspects of Earnings Mobility. Econometrica, 46(5), 985–1012.
Lokshin, Michael and M. Ravallion (2001), "Nonlinear Household Income Dynamics in Transition Economies," mimeo, Development Research Group, World Bank.
Luseno, W.K., McPeak, J., Barrett, C.B., Gebru, G. & Little, P.D. (2003). The Value of Climate Forecast Information for Pastoralists: Evidence from Southern Ethiopia and Northern Kenya, *World Development*, 31(9), 1477-1494.
Lybbert, Travis, Christopher B. Barrett, John McPeak and Winnie K. Luseno (2007). Bayesian herders: asymmetric updating of rainfall beliefs in response to external forecasts, *World Development* 35(3): 480-497.
MaCurdy, T. (1982). The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis. *Journal of Econometrics* 18(1): 83-114.
MaCurdy, T. (1983). 'A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty, *International Economic Review*, vol 24, 265-289.
Manski, C.F. (2004). Measuring Expectations. *Econometrica*, Vol. 72, No. 5 (September, 2004), 1329–1376.

McKenzie, D., J.Gibbson, S. Stillman (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abraod?. *Journal of Development Economics*, 102: 116-27, 2013.
Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica*, 72 (1), 1-32.
Naschgold, F. (2012). Poor stays poor – household asset poverty traps in rural semi-arid India. Cornell University.
Santos, P., and C. B. Barrett. (2006). Informal insurance in the presence of poverty traps: Evidence from southern Ethiopia, Cornell UniversityWalley, P. (1991). Statistical reasoning with Imprecise Probabilities. Chapman and Hall, 1991.

8 Appendix

Appendix A - Eliciation of subjective expectations, survey instruments

To elicit subjective expectations about future income, the respondent was asked to think about a very positive and a very negative scenario about next year income and report the maximum possible income and the worst possible income. The specific questions asked were:[19]

Minimum: *Imagine that you have **a very good year**, every member of working age in the household managed to have work, and there were no droughts or anything the like. What would be the **maximum** amount of income your household would receive in such a situation in one year?*

Maximum: *Now imagine the total opposite: the harvest is bad; animals get sick, finding work is not possible. What would be the monthly income of your household in such a situation?*

Once the range of variation has been established, the interviewer divides the resulting interval in four equal intervals by identifying cut off points A,B and C. She then asks the following questions:

*How likely do you think it is that your income in the coming year will be **higher** than _________ (**A/B/C**) Rupees?*

The concept opf probabilities and probability laws

As the majority of respondents had no or very little schooling it could not be expected for them to know the concept of probabilities and the probability laws that these follow. A visual aid – namely a ruler - was employed to help the respondents. A short introduction as to how to answer such a question was given to the respondents in the following form.[20]

> *We have here a ruler with a scale from 0 to 100. We will use this as an **indicator of how sure you are that a situation will happen in the future**. Let's take rain as an example: How sure are you that it will rain sometime tomorrow?*
> *1. If you are absolutely sure that it will rain, point to the 100 on the ruler.*

[19] Alternatively, one could have chosen the same range for all households based on secondary data. We decided against this option so as to reduce the problem of anchoring. See Tversky & Kahneman (1974) for a discussion.

[20] This was done after asking about the minimum and maximum expected income and before eliciting the probabilities of the thresholds (A, B, C) occurring.

2. If you are absolutely sure that it will not rain tomorrow, point to 0 on the ruler.
3. If you are not sure whether it will rain or not but think that it is more likely to rain than not, point somewhere on the ruler between 0 and 100 but closer to 100 than to 0.
4. If you are not sure whether it will rain or not but think it is more likely that it will not rain, point somewhere on the ruler, but closer to 0 than to 100.

Subsequently, respondents were asked to give their belief on the probability of rain the coming day.[21]

The understanding of the concept of probabilities is one important factor in the elicitation process; a second one is the understanding of certain basic probability laws. Important in this context is the concept of monotonicity. Since the income thresholds A, B, and C are increasing, the probability of earning exceeding these thresholds should not increase. In order for respondents to grasp this concept, respondents were not only asked about the probability that it would rain tomorrow but also that it would rain within the coming week and within the coming month. The probability of these occurrences should not decrease for monotonicity to hold.

Appendix B - Realization and Expectation over the same period

We discuss here one caveat in the data, which is that the second survey round was conducted not exactly one year after the first. This implies that the recorded expectations and realizations in the first survey do not cover the same time period as expectations and realizations recorded in the second survey round. There is on average a 4 months (117 days) delay between the first and the second round interview as shown in Table 12.[22] We look at the time overlap of the expectation data from the first survey round and the realization data from the second survey round to explain, why we do not think this is a major problem.

Table 12: Number of days that 2nd interview was more than one year after the first interview

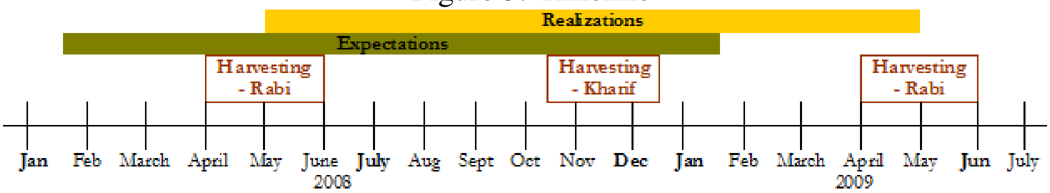
	obs.	min	max	median	mean	std.dev.
Number of days	900	50	177	114	117	22

[21] Although the survey was conducted in one of the most drought-prone areas in India and outside the monsoon season, it was possible to get variation in the responses to questions on the probability of rain. During the time the survey was conducted, it rained on several days – contrary to what is typical in the area.

[22] This discrepancy happened due to funding confirmation having been later than expected.

A time-line is presented in Figure 3, which starts with January/February 2008, the time of the first survey round interviews and ends with July 2009, the date of the final interview during the second survey round. The figure also displays the approximate time periods of the harvesting seasons for the two main cropping seasons in the area, namely Rabi and Kharif.

Figure 3: Timeline



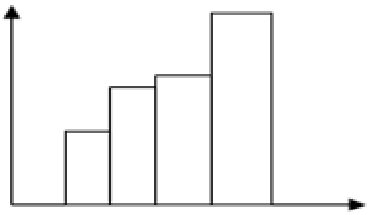
We look at these as most households derive their main income from agricultural activities, as already explained before. We can see that the expectation period included Rabi and Kharif of 2008, while the realization period includes fully the Karif season 2008 but overlaps partly with the Rabi season 2008 and 2009. We believe it reasonable to assume that households rather included profit from the 2008 Rabi season when reporting their realizations than from 2009 since selling of the produce would happen after the harvest. Based on this assumption the main income periods fall within the overlap of Realizations and Expectations and make a comparison meaningful.

Appendix C

Table 13: Probabilities assigned to sections of income distribution (2009)

	obs.	min	max	median	mean	std.dev.
Min - LQ	919	0	60	10	16.14	10.40
LQ - Midpoint	919	5	50	20	22.24	10.02
Midpoint - UQ	919	4	60	20	23.17	9.50
UQ - Max	919	0	76	40	38.45	15.56

Figure 4: The piecewise uniform distribution (2009). Household income



Appendix D

Table 14: Summary statistics of income variables (2008 and 2009) - level

	Level					
	2008 - Round 1					
Variable	Obs	min	max	p50	mean	sd
Normal income	1,029	1,200	850,000	50,000	67,001	67,007
Last year's income	1,035	3,300	660,000	54,000	71,529	63,974
Expected Income:						
Mean	986	-404,082	2,804,990	107,999	185,719	288,144
Standard deviation	986	1,297	3,798,965	48,889	127,303	280,103
Coeff. of Variation	982	-3.183	3.490	0.460	0.491	0.319
Min	1,023	2,000	350,000	40,000	53,461	44,889
Max	1,025	3,000	600,000	70,000	87,604	67,415
	2009 - Round 2					
Variable	Obs	min	max	p50	mean	sd
Normal income	946	3,900	520,000	60,700	69,781	45,237
Last year's income	946	3,900	588,800	60,100	69,672	47,823
Expected Income:						
Mean	904	7,087	777,499	92,499	115,455	90,080
Standard deviation	905	1,228	496,772	33,123	50,955	58,018
Coeff. of Variation	898	0.141	0.784	0.369	0.388	0.136
Min	942	3,000	410,000	40,000	47,883	37,230
Max	946	4,000	500,000	73,500	80,933	51,993