Forecasting when it matters: Evidence from Semi-Arid India

Xavier Giné* Robert M. Townsend James Vickery

World Bank MIT FRB of NY

First draft: June 11, 2007

This draft: August 28, 2009

Abstract

In regions of India where cultivation is rainfed, the optimal time to plant coincides with the onset of the monsoon. In this paper, we elicit from farmers their prior beliefs about the timing of the monsoon and assess their accuracy by comparing them to historical data. We find substantial heterogeneity in beliefs and

*Financial support from CRMG, GARP and World Bank is gratefully acknowledged. We want to thank Kumar Acharya, the survey team at ICRISAT and specially Dr. K.P.C Rao for many discussions about agricultural practices in semi-arid India and for their efforts in collecting the survey data. Dr. Piara Singh from ICRISAT provided valuable help in the computation rainfall absorption of different soil types. We also thank Jean Marie Baland, Shawn Cole, Adelaine Delavande, Quy-Toan Do, Pascaline Dupas, Michael Kremer, Lance Lochner, Hanan Jacoby, David Mckenzie and participants in several conferences and seminars for valuable comments. Paola de Baldomero Zazo, Helene Bie Lilleor, Francesca de Nicola and Joan Pina Martí provided outstanding research assistance. The views expressed in this paper are the authors' and should not be attributed to the World Bank, Federal Reserve Bank of New York or the Federal Reserve System. Email addresses: xgine@worldbank.org, rtownsen@mit.edu, and james.vickery@ny.frb.org.

in accuracy. More importantly, elicited beliefs can explain differences in observed behavior. We then investigate the sources of inaccuracy and find that they fit well the predictions of a simple model of costly acquisition of information: farmers that have less access to risk coping mechanisms have more accurate priors. Finally, we show that accuracy leads to average income gains of 8 to 9 percent of agricultural production, suggesting that mistakes in the timing of planting can be costly.

Keywords: Information Acquisition, Expectations, Climate.

JEL Codes: D81, D84, O13, Q54.

Uttara chusi, yattara gampa.

Wait for the Uttara rains; if they don't come, leave the place.

Telugu Proverb

1 Introduction

Weather risk is a major source of income fluctuations for rural households in developing countries. Rosenzweig and Binswanger (1993), for example, find that the delay of the monsoon in semi-arid India can have considerable negative effects on agricultural yield and profits. If the monsoon were to arrive one standard deviation late, the poorest quartile of the households in their data would experience a reduction of 35 percent in agricultural profits, while for the median household, the drop would be of 15 percent.

With complete and frictionless financial markets, households would be able to protect consumption from weather shocks fairly well. But because formal insurance markets in developing countries are typically missing, households have to rely on the ex-ante and ex-post risk coping strategies that typically trade expected profits for lower risk (Walker and Ryan, 1990).

One such strategy, particularly if agricultural production is rainfed, is to choose an optimal sowing window (Fein and Stephens, 1987; Rao et al., 2000 and Gadgil et al., 2002). Farmers in semi-arid India, where the main growing season runs from June to November (coinciding with the monsoon), wait for the onset of the monsoon to start planting. If planting occurs when the first rains come, but these rains are scattered and fall several days apart, the seeds may not germinate. In order to salvage the production, farmers are forced to replant, but they may decide to abandon the crop altogether. Either strategy results in significant losses. On the other hand, being too conservative by postponing planting until one is certain that the monsoon has arrived is also costly, because yield will typically be lower (Fafchamps, 1993; Rao et al., 2000 and Singh et al., 1994). In short, when the first rains of the season come, farmers must assess whether they are just early pre-monsoon rains, in which case they should postpone planting, or whether the rains signal the onset of the monsoon, in which case they should plant immediately.

In 2006, about a quarter of our sample of 1,054 farmers, interviewed in 2004 and 2006, had replanted in the past and a full 73 percent had abandoned the crop at least once due to lack of rain. The extra expenses born by those that replanted account for 20 percent of total production expenses. This evidence suggests that mistakes about the precise timing of the monsoon are costly. Thus, it seems that farmers would benefit greatly from having accurate priors of the onset of the monsoon.

Of course, by the time the planting decision is made, farmers may have received numerous signals about the likelihood of the monsoon onset. But because these signals are not conclusive, the *prior* distribution still matters as it affects the farmers' *conditional* expectation.

In this paper, we elicit the farmers priors about the timing of the monsoon and study whether they share the same beliefs, and if not, determine which farmers are the most accurate and why. The survey covers 1,054 farming households living in 37 villages of two districts in Andhra Pradesh, India. Because the survey was conducted as part of an evaluation of a weather insurance pilot, sampled villages are located less than 10 miles away from the nearest weather station. Experimental methods were used to first elicit the respondent's subjective definition of the onset of the monsoon, and then its subjective calendar distribution. To elicit the subjective definition, each farmer reported the minimum depth of soil moisture that he would require to start planting. This measure was then converted, using the absorption capacity of the respondent's specific soil type, into a quantum of rainfall. The subjective calendar distribution was obtained by giving each respondent 10 stones and a sheet of paper with boxes corresponding to 13-14 day astral periods called "kartis" (or naksatra in Sanskrit) based on the traditional solar calendar (Fein and Stephens, 1987; Gadgil et al., 2002). The respondent was then instructed to place the stones in the different boxes according to the likelihood that the monsoon would start in the period indicated by each box. We then compute for each respondent a chi-square statistic that compares the elicited subjective distribution to the corresponding historical one constructed with the respondent's own definition of the onset of the monsoon and the historical rainfall data from the nearby rainfall gauge.

We show that this elicitation method delivers informative prior subjective expectations and that individuals behave according to them. First, in places where rains are more erratic, farmers require a significantly larger quantum of rainfall (or depth of soil moisture) to start planting, which is consistent with the prediction of a simple model of irreversible investments where signals of differing quality are received prior to the investment decision (planting in our case). When the signal about the true state of nature is of good quality, that is, when the first rains are informative about whether the monsoon has indeed started, the minimum rainfall that the farmer will require to plant will be lower (and the percentage of expenditures before the monsoon higher). Second, farmers who believe the monsoon will start later are also more likely to plant later. Likewise, they are less likely to replant, have purchased a lower share of total production inputs before the onset of the monsoon and are more likely to buy weather insurance, since according to their beliefs, the probability of a payout is higher. All of these findings provide strong evidence that individuals make decisions according to their prior expectations, even when controlling for self-reported proxies of risk aversion, discount rates and the actual start of the monsoon. To the extent that differences in behavior can be explained by differences in expectations independently of differences in parameters of the utility function, eliciting expectations seems to be warranted.

We also show that the accuracy of farmers' prior expectations seem to be dictated by a simple cost-benefit analysis. Accurate farmers are poorer, tend to have more rainfall dependent income and report being credit constrained. In sum, accurate farmers are less able to cope with weather risk and have income sources that depend more heavily on it. Thus, accuracy is explained rationally by how relevant the event is to the forecaster, rather than by heuristics or "rules of thumb" developed in the psychology and behavioral economics literature (Kahneman, Slovic and Tversky, 1982; Rabin, 1998 or, more recently, DellaVigna, 2007 for a review).

Finally, we show that inaccuracy about the onset of the monsoon is costly. In particular, we find that an improvement in our measure of accuracy from the 25th percentile of the distribution to the 75th percentile would increase gross agricultural income by 8 to 9 percent. Related, we also show that the accuracy is correlated with a lower probability of replanting, suggesting that accurate farmers make less planting mistakes.

This paper also contributes to a growing literature that measures expectations on various outcomes (see Manski, 2004 for an excellent review and Norris and Kramer, 1990 for an early review of elicitation methods applied to agricultural economics. More recently, Delavande et al. (2009) and Attanasio (2009) also review the literature in developing countries and advocate for collecting expectations data). In developing countries and using similar elicitation methods to the one used in this paper, Luseno et al. (2003) and Lybbert et al. (2005) study the extent to which cattle herders in Kenya update their priors on rainfall expectations in response to new information. They find significant updating but little change in behavior after the updating, possibly due to the flexible nature of their income generating activities with respect to weather changes, as herds can be moved to greener pastures. Delavande and Kohler (2007) elicit expectations of HIV-infection risks in rural Malawi. These studies, like ours, elicit expectations about events in which respondents may have had substantial experience. In contrast, McKenzie et al. (2007) finds sizeable inaccuracies when individuals lack experience about the event. First time intending migrants tend to under-estimate the prospects of finding a job abroad and the incomes they could earn, when compared to the actual experience of a valid group of migrants. Because we study an event for which objective historical data exists, in this paper we focus on differences in accuracy across individuals rather than differences across events with differing prior knowledge. In addition, unlike the previous literature, we are able to estimate the cost of inaccuracy.

The rest of the paper is organized as follows. The next section describes the context and the data collected. Section 3 presents a simple model that will guide the empirical strategy of Section 4, and finally Section 5 concludes.

2 Data and Context

This paper uses two sources of data: survey data and historical rainfall data. Survey data were collected for the baseline of an impact evaluation of the weather insurance pilot described in Giné, Townsend and Vickery (2008). The survey took place after the 2004 harvest, and covered 1,073 households in 37 villages from two drought prone and relatively poor districts in the state of Andhra Pradesh, India. We resurveyed the same households in 2006, so a few variables are constructed using the follow-up survey. Figure 1 shows the location of Mahbubnagar and Anantapur, the two districts in AP where the survey was conducted.

The sampling framework used a stratification based on whether the household attended a marketing meeting and whether the household ended up purchasing a deficit rainfall insurance policy in 2004. All summary statistics report weighted data (again see Giné, Townsend and Vickery, 2008 for further details). Throughout the analysis, if "purchase of insurance" is not the dependent variable, the stratification variables (attended meeting and purchasing insurance) are always included as controls. When "purchased insurance" is used as the dependent variable, we follow Manski and Lerman (1977) and use a weighted regression.[1]

[1]Except for the case when "purchase of insurance" is the dependent variable, either weighted or unweighted regressions would provide consistent estimates of model parameters. In the analysis, we include the stratification variables as controls rather than using a weighted regression because this typically results in a loss in efficiency. In all regressions, the stratification variables are insignificant. See DuMouchel and Duncan (1983) for details.

All survey villages are located less than 10 miles away from the nearest mandal (county) rainfall gauge. Figure 2 shows with a dot the location of the 23 villages in Mahbubnagar. In addition, the location of the five rainfall stations is also shown along with a 10 mile radius circumference. As can be seen, villages are close to rainfall gauges thus minimizing the difference between rainfall observed in the gauge and in households' plots.

Villages are assigned to the closest rainfall station. Table 1 presents the number of villages and households assigned to it, the average distance from each gauge to the villages assigned to it and the number of years with available daily data. There are five rainfall stations in Mahbubnagar district (Panel A) and five in Anantapur district (Panel B).

The main cropping season runs from June to November and is mostly rainfed.[2] Farmers grow a variety of cash and subsistence crops that vary in their yield's sensitivity to drought. The main cash crops grown in the area are castor which is mostly grown in Mahbubnagar and covers 34 percent of its cultivated land in the sample, groundnut, mostly grown in Anantapur and covering half of the cultivated land in the Anantapur sample, and paddy (close to 9 percent of the combined cultivated land). Paddy is almost exclusively irrigated (84 percent of all plots use irrigation).[3] The subsistence crops grown in the area are grams (redgram and to a lesser extent greengram, covering 10.6 percent of all cultivated land) and sorghum (6.7 percent). All subsistence crops as well as castor and groundnut are typically rainfed (less than 5 percent of plots are irrigated).

When irrigation cannot be used to smooth the vagaries of the monsoon, farmers try to choose an optimal sowing window (Fein and Stephens, 1987; Rao et al., 2000; Gadgil et al., 2002). When the first rains of the season come, farmers must assess whether these are just early pre-monsoon rains or whether the rains signal the actual onset of the monsoon.

[2] Although wealthier farmers have irrigation equipment, irrigation is mostly used for the shorter growing season during the dry months (January-May).

[3] Maize, sunflower are cotton are also grown but to a lesser extent, since their combined area is less than 10 percent.

In the former case, farmers should postpone planting because if they were to plant, the seeds would not germinate and they would either be forced to replant or abandon the crop. In the latter case, they should plant immediately because if they were to wait, the seeds would not capitalize on the rain already fallen, thus undermining its yield (Fafchamps, 1993; Rao et al., 2000; Singh et al. 1994).

Table 2 presents suggestive evidence of this inverted U-shape relationship between yields and the timing of planting by regressing the cross-section of reported yield per acre in 2004 of paddy and the main rainfed crops (castor, groundnut, sorghum and redgram) against a linear and quadratic term of the time when the crop was planted. We use kartis instead of weeks or months because it is the unit of time that farmers typically use (Fein and Stephens, 1987; Rao et al., 2000; Gadgil et al., 2002). Table 3 reports the relevant kartis for the main cropping season assuming that the first kartis of the calendar year take value 1, the second value 2 and so on. The regression in Table 2 includes as controls crop times district dummies, plot level characteristics and household fixed effects. The variables of interest are kartis interacted with a crop and district dummy. As discussed above, for the rainfed crops (that is, excluding paddy), the p-value of an F-test that all kartis times crop times district coefficients are jointly insignificant is 0.008 for Anantapur and 0.025 for Mahbubnagar. The corresponding p-values for the kartis squared are 0.003 and 0.019, respectively. Thus, for the rainfed crops, the linear and quadratic terms of the time of planting are jointly significant and of the expected sign.

In 2006, we also collected information about replanting decisions. Twenty-two percent of households have replanted in the last 10 years at least once, and almost three quarters have abandoned the crops altogether at some point over the same period. The main reason for abandoning the crop is failure of the seed to germinate and either lack of capital to purchase additional seeds, or low expectations of subsequent rains to warrant replanting. In 2006, roughly 3 percent of the sample replanted some crop, spending an additional Rs 5,000 (USD 86) accounting for 23 percent of total production costs.

Another cut at the data reveals large differences in the expenses incurred before the onset of the monsoon. Before the arrival of the rains, farmers make several production related expenditures to prepare the land according to the crops to be grown, apply manure and purchase seeds, especially if these are hybrid or improved. Given the weather uncertainty, if farmers waited for the rains, they could better decide what crops to plant and therefore prepare the land accordingly. However, farmers prepare the land in advance to be ready to sow when the rains come and also purchase the seeds in advance because they fear that prices might go up or availability may be difficult once widespread rains are received. In Mahbubnagar, these expenses amount to one third of all production costs, but in Anantapur they are virtually zero. We will return to this stark difference in the analysis presented in Section 4.

All in all, this evidence suggests that by having accurate priors about the start of the monsoon, farmers could avoid costly mistakes.[4]

It is important to stress that farmers decide when to plant based on their conditional expectation that the monsoon has arrived. If upon receiving a signal prior to the monsoon, farmers knew with certainty the exact timing of the onset and thus the optimal time to plant, then priors would not matter, since they would simply wait for the signal to determine when to plant. However, as evidenced by a long and rich folk tradition of trying to predict the arrival of the rains, these signals are hardly conclusive. This accumulation of indigenous knowledge over thousands of years is formalized in the literature, folk songs, and proverbs or sayings. For example, farmers use the color of the sky, the shape and color of the clouds, the direction of the winds, the appearance of certain insects or migratory birds, and so on, to update the probability that the monsoon has arrived.[5] Thus, the

[4]Although what ultimately matters for a good harvest is total accumulated rainfall at key points of the cropping season, the onset of the monsoon, as defined in the next subsection is correlated with total accumulated rainfall (p-value is 0.003). Thus, the later the onset, the lower will be accumulated rainfall during the cropping season.

[5]Besides these signals for short-term forecasting, classical astrological texts also describe a long-term

prior distribution is still important as it influences the update.

2.1 Onset of Monsoon

We use two definitions of the onset of the monsoon. The first definition uses the respondent's self-reported minimum quantum of rainfall required to start planting. The survey asked "What is the minimum amount of rainfall required to sow?" and also "What is the minimum depth of soil moisture required to sow?". Only 10 percent of farmers provided an answer when asked in millimeters, but all farmers responded using the depth of soil moisture, so we used the soil's absorption capacity to convert soil depth into millimeters of rainfall.[6] We label this quantum of rainfall from each respondent, the individual definition of the onset. The second definition is simply the average in each district of the individual definitions just described. The average onset of the monsoon in millimeters is 28.95 mm in Mahbubnagar and 33.05 mm in Anantapur.

Since a given farmer may have plots of differing soil texture, the quantum of rainfall is computed as a weighted average of a millimeter amount for each soil texture, weighted by plot size. Eighty percent of cultivable land in Anantapur is red soil, with texture either loamy sand or loam, and the rest is black soil, with texture either silty clay or clay loam. Mahbubnagar has more of a balance, with 66 percent of the soil being red. (See "Land Characteristics" in Table 5 described in Section 2.4).

and short-term forecasting method based on the so-called "pregnancy" of clouds. The idea of the long-term method is that the weather patterns (the so-called symptoms of cloud pregnancy) on November 26, are used to predict rainfall on June 9th, 195 days later. If November 26 is a bad day, but December 6 is promising, then the monsoon would start not on June 9th but on June 19th. The short-term method requires astrologers to look at the amount of rainfall in the next few days after the onset of the monsoon in order to predict the quantity of rainfall for the entire rainy season (Fein and Stephens, 1987).

[6]The calculations for rainfall penetration for various soil textures were made by Dr. P. Singh (ICRISAT) using the generic values of field capacities under the following assumptions: (i) The top soil (1-1.5 foot) is completely dry, (ii) no runoff occurs (iii) moisture is primarily determined by the texture and structure of the soil and (iv) evaporation from soil surface is ignored.

Figure 3 plots for a particular rainfall station in each district, the time that it took daily rainfall to accumulate to the definitions described above each year, measured in days since June 1st. For example, the left plot shows that according to the historical daily rainfall for the year 2000, it took 33 days for rainfall as measured in the Hindupur gauge to accumulate to 33.05 mm, the average of the individual definitions in Anantapur district. In the same year (right graph), it only took 3 days for rainfall measured in Mahbubnagar to accumulate to its district average.

Figure 3 shows that there is volatility in the onset of the monsoon, but if signals about the onset such as the Indian Meteorological Department (IMD) forecast were conclusive, then such volatility would not be of concern as farmers would just act upon the signal. Figure 4 plots the more relevant onset of the monsoon conditional on the forecast by IMD.[7] In particular, it plots the difference in days between the actual onset of the monsoon of Figure 3 in a given year and the forecast from the IMD for that year.[8] We call this difference the "conditional" start of the monsoon.

Two observations are in order. First, at least visually from Figure 3 and 4, there does not seem to be a linear or quadratic trend in the start of the monsoon over time. This is consistent with an article that received a lot of press coverage by Goswami et al. (2006) which finds that while there is no trend in the average accumulated rainfall over the season, the frequency of extreme events such as heavy rains and prolonged droughts increased over the period 1951–2000. Thus, global warming may have increased the variance of rainfall during the monsoon, but not necessarily its average nor its onset.[9]

When we regress the onset date (in days since June 1st) against a linear and quadratic

[7] The IMD issues a single forecast of the onset of the monsoon over Kerala, around May 15th. We extrapolate this forecast to other regions adding the normal onset dates plotted in Figure 1 to the forecast over Kerala.

[8] We obtain a very similar picture when we use the first 20 mm of rainfall or the first two consecutive rainy days as a signal, instead of the IMD forecast.

[9] See also Gadgil and Gadgil (2006) and Mall et al. (2006) for further evidence on all-India climatic trends and a review of their effect on agriculture.

trend using year and rainfall station observations in each district (a total 170 observation in Anantapur and 158 in Mahbubnagar), neither the linear nor the quadratic coefficients are significantly different from zero (results not reported).

The lack of a trend over time in the timing of the monsoon is important because if such a trend existed, one would probably have to weight recent years more heavily when constructing the historical distribution of the onset of the monsoon. We construct the historical distribution by computing for each rainfall station, the number of years with available data for which the monsoon arrived in a particular kartis, weighting each year equally.

Second, both the onset of the monsoon in Figure 3 and the conditional onset of Figure 4 are more volatile in Anantapur than in Mahbubnagar. Indeed, Figure 5 compares the cumulative density function of the conditional onset of the monsoon arbitrarily defined as 20 mm (left panel) or 40 mm (right panel), measured in kartis. In both cases, Anantapur clearly dominates in a second-order stochastic sense Mahbubnagar. In the next section, we will assess how farmers in Anantapur respond to this more erratic environment.

2.2 Consistency of Subjective Distributions

As explained in the introduction, the subjective distribution was obtained experimentally by giving the respondent 10 stones and a sheet of paper with boxes corresponding to the different kartis. The respondent was then instructed to place all the stones in the different boxes according to the likelihood that the monsoon would start in each period. Thus, the question did not specify the following year's monsoon, but rather the monsoon in general, and should be interpreted as the respondent's prior distribution of the onset of the monsoon. This method of elicitation has been successfully used in Barrett et al. (2006) and Delavande and Kohler (2007), among others (see Delavande et al. 2009 for a review). By giving stones, one ensured that probabilities added up. It is worth noting that every respondent provided a subjective distribution and that no respondent allocated less than the total 10 stones given. In addition, no respondent allocated a stone in the first or last kartis of the predefined support, indicating that it was sufficiently wide to not constrain the respondents' answers. One concern is that respondents were constrained by the limited number of stones given, especially in Anantapur where there is more uncertainty about the start of the monsoon. If that were true, then the predictions in Anantapur would be less accurate. However, as we show below, respondents in Anantapur predict more accurately than in Mahbubnagar.

Table 4 provides further evidence of the consistency of subjective distributions as compared to the historical ones, which are computed under both definitions: the individual self-reported minimum quantum of rainfall required to plant ("Individual") and the district average of these individual definitions ("District"). Observations are weighted according to the stratification weights. Under "Support of the distributions", we can see that the lower and upper bound of both the subjective and historical distributions are remarkably similar, except perhaps in Anantapur, where respondents seem to underestimate the range. When we look at the percentage of respondents that used different number of bins, we find that nobody put all the 10 stones in a single bin. Most respondents allocated stones in 3 or 4 different bins. In "Moments and Properties of Distributions", we see that in Mahbubnagar respondents under-predict the mean and over-predict the variance, while the reverse is true in Anantapur.

2.3 Accuracy

We measure the accuracy of the subjective expectations by computing a simple chi-square statistic for binned data. Let H_k be the number of years of available data when the monsoon arrived in kartis k as recorded in the rainfall station assigned to the respondent, and S_k the number of stones from the available ten stones, that the respondent placed in the same kartis k. Because the number of years is different from the number of stones, the chi-square statistic is

$$\chi^2 = \sum_k \frac{(\sqrt{S/H}H_k - \sqrt{H/S}S_k)^2}{H_k + S_k}, \tag{1}$$

where

$$S = \sum_k S_k \qquad \text{and} \qquad H = \sum_k H_k.$$

This statistic follows a chi-square distribution with degrees of freedom given by the number of bins (i.e. kartis in the sheet of paper given the respondent) minus 1. Although the terms in the sum of expression (1) are not individually normal, because both the number of bins and the number of events in each bin is large, the chi-square probability function is a good approximation to the distribution of (1).[10,11]

Under "Accuracy", Table 4 reports the percentage (appropriately weighted) of individuals in each district whose subjective distribution is statistically different from the historical ones at different significance levels.

Accuracy is lower when the individual definition, compared to its average across respondents in the district, is used. This suggests that inaccurate forecasters miss both the timing of the onset of monsoon as well as its definition. When the district average is considered, mistakes across respondents are compensated in such a way that it the

[10]See Press (1992) for further details and references.

[11]The chi-square statistic above assumes implicitly that the data come from two random samples. For the historical data, this is certainly the case because we only have available a sample of years. For the subjective data, we assume that each stone corresponds to a random draw from the respondent's subjective distribution. Alternatively, we could have computed the chi-square distribution assuming that the subjective distribution was the true distribution. In this case, the chi-square statistic would be

$$\chi^2 = \sum_k \frac{(H_k - s_k)^2}{s_k}$$

where now s_k is the expected number of years that the monsoon arrives in kartis k. The problem with this approach is that if $H_k > 0$ but $s_k = 0$, then the χ^2 statistic is infinity and thus intractable in the analysis of Section 4.

respondents' subjective distribution of the timing of the onset is closer to the historical distribution, thus lowering the chi-square statistic on average.

All in all, 40 percent in Mahbubnagar but only 20 percent in Anantapur report a subjective distribution that is significantly different at the 10% level from the historical one computed using the individual subjective definition. These numbers are 23 and 14 when the district average of individual definitions is used. It appears that respondents are more accurate in Anantapur where the onset is more volatile, and thus where it matters more, but we will explore in more detail these differences in accuracy in Section 4.

2.4 Summary statistics

Table 5 presents some summary statistics for the household variables used in the analysis. The Appendix contains a detailed description of how these variables are constructed. All statistics in Table 5 are weighted by the sampling weights described in Section 2.

We first compute two variables that represent parameters of the household head's utility function. These parameters were estimated experimentally through a series of hypothetical scenarios presented as part of the survey. The variable "risk aversion" is constructed based on questions where the household head chooses between a series of hypothetical gambles indexed by increasing risk and return (see Binswanger, 1980, 1981 and Binswanger and Sillers, 1984). Households who chose the safe bet (Rs. 50 with certainty) over any of the risky gambles were identified as risk averse and assigned a value of 1 (risk aversion = 1). Other households were assigned a value of 0 for this variable. The variable "discount rate" is measured from the elicited amount that a household head would receive today in order to be indifferent relative to a fixed amount promised in one month's time. The average for this variable is around 30 percent, suggesting a high discount factor for the households in the sample.

Next, we compute two variables that measure the extent to which farmers talk to each other about weather events privately or in public forums, such as during meetings of the Borewell User Association (BUA). We also include the age of the eldest household member, since the household head could benefit from his or her experience when forming expectations about the onset of the monsoon.[12]

We construct wealth measures that confirm that the sample consists of poor and middle-income smallholder farmers. The average value of the land owned was Rs. 246,000 (USD 5,234) and the value of the households' primary dwelling averaged Rs 69,000 (USD 1,468).

We then construct three variables that proxy for the ability that the household has to smooth income shocks. First, participation in chit funds (roscas), since these financial arrangements can be used not only for investments but also to smooth consumption (see Klonner, 2007). Second, a variable indicating whether the household was credit constrained, which is a dummy variable equal to 1 if the household applied for credit but was denied, or the household cited "no creditworthiness" or "no access to lender" as reasons why it did not apply for credit. Finally, we include total income encompassing both farming and non-farming income.

Next, we construct two variables that measure the household's exposure to weather shocks. First, its total cultivated land, and then the percentage of land devoted to paddy which requires irrigation.

Table 5 also includes the percentage of land with different textures, the prevalence of intercropping and the slope and depth of the plots.[13] The bottom of Table 5 reports basic cropping patterns followed by farmers in our sample. Interestingly, most farmers tend

[12] See Rosenzweig and Wolpin (1985) for an interesting theory that explains the predominance of intergenerational family extension, kin labor and the scarcity of land sales from the fact that elders have specific experiential knowledge about farming practices.

[13] Intercropping is a frequently used agricultural practice in the semi-arid tropics. It consists of growing two (or more crops) in the same plot planted at the same time, alternating one or more rows of the first crop with one or more rows of the second, possibly in different proportions. One crop will typically be shallow rooted, while the other will be long-rooted, ensuring that crops do not compete for soil nutrients at the same depth.

to plant all the rain sensitive crops at once. Therefore, they do not exploit a strategy of planting different plots at different times to cope with weather risk. In anecdotal conversations with farmers, we learned that they do not follow this strategy because of the relatively high fixed costs of land preparation and other inherent indivisibilities in agricultural production.

Finally, farmers in Anantapur tend to plant later, closer to kartis Punarvasu (July 6) while in Mahbubnagar they plant by June 15th. Thus, farmers in Anantapur plant much later than June 3rd, which is the normal onset of the monsoon according to the Indian Meteorological Department (IMD), drawn in grey in Figure 1.[14] The model in the following section provides a rational explanation for this apparent behavior in Anantapur.

3 A Simple Model

Given the discussion above, we consider a risk-neutral farmer that is deciding when to plant a rain-sensitive crop after observing the first rains of the cropping season. To simplify matters, we assume that there are two periods when the farmer can plant. In period 1, before the decision to plant is made, the farmer observes a signal about the true state of nature in period 2 that will be used to update her prior. Planting is irreversible, in the sense that if the farmer decides to plant in period 1 then it cannot do so in period 2.

In what follows, we first characterize the optimal planting strategy as a function of the accuracy of the signal and later, we introduce the possibility of devoting resources (attention) to increasing the accuracy of the prior in an attempt to minimize planting

[14]Murphy and Winkler (1984) describe the field of meteorology as one in which probability forecasts are very accurate, possibly due to the wealth of available data and the long tradition in forecasting. In contrast, this finding suggest that the IMD forecast of the onset of the monsoon is less accurate than the farmer's expectations, perhaps because the focus of IMD's forecast is unrelated to the optimal sowing time that farmers in a specific region should follow.

mistakes.

3.1 Optimal Planting Strategy

The basic tradeoff in the planting decision is as follows. Planting in period 1 maximizes accumulated rainfall needed for crop growth, but if it doesn't rain in period 2, the seeds planted in period 1 may not germinate resulting in crop failure. Alternatively, if the decision to plant is postponed until period 2, then the farmer can make a more informed decision, but it may be too late as the crop will not benefit from the rainfall in the second period resulting in a lower yield. Thus, because planting is an irreversible investment, mistakes are costly.

We model this tradeoff as follows. As mentioned, in period 1 the farmer observes an amount of rainfall r which can take on three values, r_H (high), r_M (medium) and r_L, (low), where $r_H > r_M > r_L = 0$. This amount is *informative* of the realized state of nature s in period 2, which can be H for high rains, M for medium rains and L for low rains.[15] If the farmer plants in period 1, then she obtains income y_s, $y_H > y_M > y_L = 0$ depending on the state of nature realized. If, on the other hand, the farmer decides to wait until period 2 to plant, she obtains income αy_s, $s = H, M$, where $\alpha \in (0, 1)$ to reflect the cost of waiting. Production costs are $c > 0$, and we assume that $\alpha y_M > c$, so that it is optimal to plant if the state is H or M, even if the farmer waits to plant in period 2.

The probability of observing signal r conditional on the state of nature s is given by $\Pr(r_s|s) = \frac{1+2\delta}{3}$ and $\Pr(r_{\neg s}|s) = \frac{1-\delta}{3}$, $s, \neg s = H, M$ or L.

Letting $\Pr(s) = \mu_s$ be the farmer's prior of state s occuring in period 2, we can write the unconditional probability of receiving signal r_H as

$$\Pr(r_H) = \frac{1+2\delta}{3}\mu_H + \left(\frac{1-\delta}{3}\right)(1-\mu_H).$$

Using Bayes Law, we can write the posterior probability of the state of nature being

[15] See Hirshleifer and Riley (1995) or Gollier (2006).

H conditional on observing signal r_H as

$$\Pr(s = H|r_H) = \frac{\Pr(r_H|s = H)\Pr(s = H)}{\Pr(r_H)} = \frac{1}{1 + \frac{1-\delta}{1+2\delta}\frac{1-\mu_H}{\mu_H}}.$$

One can derive analogous expressions for $\Pr(s = M|r_M)$ and $\Pr(s = L|r_L)$. The decision to plant will depend on the priors μ_s, $s = H, M, L$ and δ, which measures the degree to which signal r is informative.

When $\delta = 0$, signal r is uninformative and priors are not updated $\Pr(r_s) = \mu_s$, $s = H, M, L$. However, as δ increases, so does the ex-post ability to assess the rains in period 2.

Now define Δ_H as the difference in expected utility between planting and waiting in period 1 when signal r_H is received: it is optimal to plant if $\Delta_H > 0$ and to wait otherwise. Using the conditional probabilities defined above, we can write Δ_H as a function of δ as

$$\Delta_H(\delta) = \frac{1}{3\Pr(r_H)} \left[(1 - \alpha)\left[(1 + 2\delta)\mu_H y_H + (1 - \delta)\mu_M y_M\right] - (1 - \delta)\mu_L(c - y_L)\right].$$

Defining Δ_M and Δ_L analogously, it is easy to show that when the signal is uninformative ($\delta = 0$), then

$$\Delta_H(0) = \Delta_M(0) = \Delta_L(0).$$

To make things interesting, we assume that $\Delta_s(0) < 0$ for $s = H, M$ and L, so that in this case it is always optimal to wait. One can also show that Δ_H and Δ_M are increasing in signal accuracy δ, whereas Δ_L is decreasing. If we further assume that $\mu_H y_H > \mu_M y_M$, then $\Delta_H(\delta) > \Delta_M(\delta)$ for $\delta \in (0, 1]$.

Figure 6 plots the functions Δ_s functions assuming equal priors.[16] From Figure 6, we see that the optimal planting strategy for the farmer when the signal is r_H (r_M) is to plant, as long as $\delta \geq \delta_H$ ($\delta \geq \delta_M$). If r_L is received, the farmer should wait. Formally, the cutoff values δ_H and δ_M are defined by $\Delta_H(\delta_H) = 0$ and $\Delta_M(\delta_M) = 0$ and can be written

[16]When $\mu_H = \mu_M = \mu_L = 1/3$), the functions Δ_s are linear in δ as in Figure 6. When other priors are assumed, then Δ_s can be concave or convex in δ, depending on the parameters.

as

$$\delta_H = \frac{\mu_L c - (1-\alpha)(\mu_H y_H + \mu_M y_M)}{\mu_L c + (1-\alpha)(2\mu_M y_M - \mu_H y_H)} \qquad \delta_M = \frac{\mu_L c - (1-\alpha)(\mu_H y_H + \mu_M y_M)}{\mu_L c + (1-\alpha)(2\mu_H y_H - \mu_M y_M)} \tag{2}$$

Thus, signal accuracy δ play an important role in the optimal strategy, as stated in the following proposition.

Proposition 1 *The amount of rain required to start planting is (weakly) decreasing in signal accuracy.*

In other words, when r_M is received, a farmer will wait if $\delta < \delta_M$, but will plant if δ is high enough, that is, $\delta \leq \delta_M$.

In addition, since the cutoff values δ_H and δ_M are both decreasing in μ_H and μ_M, the model delivers the following prediction that will be tested in the following section.

Proposition 2 *The probability of planting in period 1 is (weakly) increasing in the prior distribution μ_H and μ_M.*

For a given signal accuracy δ, a farmer that has a higher prior that state H will occur in period 2 is also more likely to plant since δ_H will be lower resulting in a higher probability that $\delta > \delta_H$.

3.2 Accuracy of Priors

Suppose now that inital priors, denoted μ_s^0, $s = H, M, L$ are inaccurate. To keep things simple, assume that $\mu_s^0 = \mu_s^* + \epsilon$, for $s = H$ and $s = M$ and $\mu_L^0 = \mu_L^* - 2\epsilon$, where μ_s^*, $s = H, M, L$ are the true priors. In addition, suppose that farmers are endowed with one unit of attention a that can be devoted to improving the accuracy of the prior or to another activity whose return is uncorrelated with the monsoon, like growing an irrigated crop. In particular, we assume that

$$\mu_s = a\mu_s^* + (1-a)\mu_s^0, s = H, M, L$$

so that

$$\mu_s = \mu_s^* + (1-a)\epsilon, \quad s = H, M \quad \text{and} \quad \mu_L = \mu_L^* - 2(1-a)\epsilon.$$

We can now write the expected net income from the rain-fed crop as a function of attention a and signal accuracy δ as follows

$$\begin{aligned}
\Pi_R(a,\delta) &= \Pr(r_H)\left[\Pr(\delta \geq \delta_H) EY(\mathrm{P}|r_H) + (1 - \Pr(\delta \geq \delta_H))\, EY(\mathrm{W}|r_H)\right] \\
&+ \Pr(r_M)\left[\Pr(\delta \geq \delta_M) EY(\mathrm{P}|r_M) + (1 - \Pr(\delta \geq \delta_M))\, EY(|r_M)\right] + \Pr(r_L) EY(\mathrm{W}|r_L),
\end{aligned}$$

where

$$EY(\mathrm{P}|r_H) = \frac{1}{3\Pr(r_H)}\left[(1+2\delta)\mu_H^*(y_H - c) + (1-\delta)(\mu_M^*(y_M - c) - \mu_L^* c)\right],$$

$$EY(\mathrm{W}|r_H) = \frac{1}{3\Pr(r_H)}\left[(1+2\delta)\mu_H^*(\alpha y_H - c) + (1-\delta)\mu_M^*(\alpha y_M - c)\right],$$

and analogous expressions can be derived for $EY(\mathrm{P}|r_M)$, $EY(\mathrm{W}|r_M)$ and $EY(\mathrm{W}|r_L)$. Finally, if we write δ_H in (2) as a function of ϵ and true priors

$$\delta_H = \frac{\delta_H^* n_H^* - \epsilon(1-a)D}{n_H^* - \epsilon(1-a)N_H}$$

where δ_H^*, is the cutoff value of accuracy δ_H using the true priors, n_H^* is the denominator of δ_H^* in (2) if again the true priors are used, $D = 2c + (1 - \alpha(y_H + y_M))$ and $N_H = 2c - (1-\alpha)(2y_H - y_M)$, we can then write the probability of planting as

$$\Pr(\delta \geq \delta_H) = \Pr\left(\epsilon \geq \frac{(\delta_H^* - \delta)n_H^*}{(1-a)(D - \delta N_H)}\right) \tag{3}$$

An analogous expression can be derived for $\Pr(\delta \geq \delta_H)$. Notice that mistakes happen if upon receiving signal r_H (r_M) the farmer planted because $\delta > \delta_H$ ($\delta > \delta_M$) when in fact she should have waited as $\delta < \delta_H^*$ ($\delta < \delta_M^*$); or if she waited $\delta < \delta_H$ ($\delta < \delta_M$) after receiving signal r_H (r_M) when she should have planted $\delta > \delta_H^*$ ($\delta > \delta_M^*$). Inspecting the expression in (3), we see that the probability of mistakenly planting ($\delta > \delta_H$ when in fact $\delta < \delta_M^*$) is decreasing in the gap $\delta_H^* - \delta$ and in attention a.

Figure 7 plots expected net income $\Pi_R(a, \delta)$ assuming that the error term ϵ follows a uniform distribution $\epsilon \sim U[-\bar{\epsilon}, \bar{\epsilon}]$.[17]. If we fix attention to $a = 1$, $\Pi_R(a, \delta)$ is a step-wise linear function along the δ-axis. When $\delta < \delta^*_H$, the optimal strategy is to wait regardless of the signal received, and thus net income does not depend on δ. When $\delta^*_H \delta < \delta^*_M$, it is optimal to plant only if signal r_H is received. In this case, higher signal accuracy results in higher expected income due to the lower probability of planting in period 1 when the farmer should have waited. Finally, when $\delta > \delta^*_M$, it is optimal to plant when signals r_H or r_M are received and to wait when r_L is received. In this case, the slope is higher than in the previous because the probability of planting in period 1 when the farmer should have waited is larger and thus the payoff to signal accuracy is higher. If we now fix attention to $a = 0$, $\Pi_R(a, \delta)$ is first convex and then step-wise linear. When the signal is uninformative $\delta = 0$, the optimal strategy is always to wait, regardless of the signal received and thus of the priors. Thus, as displayed in Figure 8, when $\delta = 0$, net income does not depend on attention a. In the other extreme, when $\delta = 1$, the signal is conclusive so priors do not matter either. In this case, as also displayed in Figure 8, net income is constant as a function of attention a. However, when the signal is informative $\delta \in (0, 1)$ there are two competing forces that shape $\Pi_R(0, \delta)$ into a U function. On the one hand, as δ increases, so does the scope for mistakes and so expected net income will decrease. More formally, the probability of planting ($\Pr(\delta > \delta_H)$ or $\Pr(\delta > \delta_M)$ depending on the signal received) is increasing in δ in the range where it is optimal to wait, since $\delta < \delta_H$ and $\delta < \delta_M$. On the other hand, expected net income is increasing in signal accuracy δ. Notice that as δ reaches 1, the probability of making a mistake is eventually zero and from then on, increases in attention do not increase expected net income. This can be seen in Figure 8 where we fix $\delta = bar\delta$ and plot $\Pi_R(a, \bar{\delta}) \equiv \pi_R(a; \bar{\delta})$ as a function of attention a. As described, except for the extreme cases of $\delta = 0$ and $\delta = 1$, expected net income is an increasing function of attention a.

[17]The maximum value that $\bar{\epsilon}$ can take is the minimum between $\frac{\delta^*_H n^*_H}{D}$ and $\frac{\mu^*_H y_H - \mu^*_M y_M}{y_H - y_M}$

Now let γ be the share of the farmer's land with access to irrigation that can be used to grow the other crop with return y_I. The farmer now needs to allocate her unit of attention optimally by solving the following problem:

$$\begin{aligned} max_a \quad & (1-\gamma)\pi_R(a;\delta) + \gamma y_I(1-a) \\ \text{s.t.} \quad & 0 \leq a \leq 1 \end{aligned}$$

Expected net income is in general concave in attention and the returns to the other crop are linear, so the above maximand is concave in attention and therefore, an interior solution may exist.[18]

Clearly, the larger the returns to the irrigated crop y_I or the larger the fraction of land with irrigation, the lower will attention a be. In other words, farmers with access to irrigated crops whose yields do not depend on the accuracy of the monsoon, will tend to have less accurate priors. We generalize this result in the next proposition and state a key implication in the following.

Proposition 3 *The less relevant the monsoon is to overall income, the less accurate will the prior be.*

A key implication of the proposition above is as follows:

Proposition 4 *When signals are inaccurate, more mistakes are made and thus, on average, the yields for rain-sensitive crops are lower.*

4 Empirical Analysis and Results

In this section we formally test the propositions derived from the model of the previous section.

[18] We could introduce concavity in attention in the $y_I(a)$ function, but in general this may not be enough to ensure an interior solution.

4.1 Timing of Planting under uncertainty

In Figure 4 described in Section 2, we showed that the conditional onset of the monsoon is more erratic in Anantapur as compared to Mahabubnagar. Proposition 1 in the previous section stated that in places with more uncertainty, or equivalently, when signals are less informative, farmers require a *higher* quantum of accumulated rainfall to start planting. Similarly, farmers should also purchase a *lower* fraction of agricultural inputs before the monsoon.

Both decisions have a degree of irreversibility. As already discussed, the downside of planting too early is that the farmer may have to either replant or abandon the crop if the seeds do not germinate. Likewise, preparing the land for a certain crop and purchasing the seeds before the monsoon arrives may be costly if given the patterns of the rains, it is best for the farmer to plant some other variety or crop that requires another type of land preparation.

To test this hypothesis of the model, we run the following regression:

$$Y_{isd} = \alpha_d D_d + \alpha_s V_s + X'_{isd}\beta + \epsilon_{isd},$$

where Y_{isd} is either the minimum amount of rainfall that farmer i in a village mapped to rainfall station s in district d requires to start planting or the percentage of purchases of agricultural inputs before the monsoon, D_d is a dummy that takes value 1 if the district is Anantapur, V_s is the standard deviation of the conditional onset of the monsoon, that is, the difference in days from the IMD forecast until the onset of monsoon (using the district average of individual definitions) across years at rainfall station s, and X_{isd} are household level characteristics. The coefficient of interest is α_s and we expect it to be positive and significant when the dependent variable is the minimum amount of rainfall and negative when the dependent variable is the percentage of expenditure before the monsoon. The coefficient α_d picks up variation that is specific to Anantapur, and is not explained by the proxy of uncertainty V_s.

Table 6 reports the results of the above specification. All regressions include land and other household controls. Columns 1 and 4 include the district dummy only, columns 2 and 5 include the standard deviation of the conditional onset of monsoon V_s and columns 3 and 6 include both the district dummy and the standard deviation of onset of monsoon.

As predicted by the model, the district dummy and the proxy for uncertainty are both positive and significant in columns 1 and 2. More importantly, when they appear together in column 3, only the proxy for uncertainly is significant, suggesting that it is capturing all the variation in the required amount of rainfall to start planting. Notice also that the coefficient on risk aversion is positive, as expected, but not precisely estimated. According to the historical data, the point estimate of uncertainty is equivalent to waiting on average 3 additional days in Anantapur to plant.

Finally, farmers in Anantapur do not seem to spend anything at all before the monsoon, indicating that they value being able to adapt the cropping patterns to the observed rainfall. Although the coefficient of the standard deviation of the onset of monsoon in column 5 is negative and significant, it loses predictive power in column 6. We therefore conclude that there is something peculiar about expenditures before the monsoon in Anantapur that cannot be explained by the uncertainty about the start of the monsoon.

4.2 Do farmers behave according to their prior expectations?

We test Proposition 2 by assessing whether respondent's behavior is consistent with his or her prior. We use the mean of the subjective distribution as a proxy for the household's prior expectation of the monsoon onset. To test the correlation between expectations and decisions, we run the following set of regressions:

$$Y_i = \alpha\mu_{Si} + \beta r_i + \delta k_i + X_i'\gamma + \epsilon_i,$$

where Y_i are one-time decisions taken by individual i, μ_{Si} is the mean in kartis of the subjective prior distribution (the expected onset of the monsoon in kartis), r_i is the individual definition of the monsoon in mms (minimum amount of rainfall required to start planting) and k_i is the actual onset of the monsoon computed using the "district" definition in odd-numbered columns or the "individual" definition in even-numbered columns.[19] This variable captures all the information available to the farmer when the planting decision is made. In addition, the vector X_i includes household characteristics that may influence decision-making, in particular risk aversion and the discount rate.

The coefficient of interest is again α. We expect it to be significant if prior expectations influence decision-making. Notice that by including the actual start of the monsoon, we are able to assess the relative importance of the unconditional vis a vis the conditional expectation. In addition, we control for the individual definition in order to isolate differences in the expected timing of the monsoon from differences in the individual definition.

We consider the following decisions as dependent variables. First, whether the household bought rainfall insurance in 2004 (in general and conditional on attending a marketing meeting).[20] The rainfall insurance paid off if accumulated rainfall until a predetermined date was below a certain trigger. Because the dates of the policy coverage were determined in advance, households that expected the monsoon to start later would also find that the insurance policy more attractive, because the elapsed time from the expected monsoon arrival until the fixed end-of-coverage date would be shorter, thereby increasing the chances of a payout. In sum, households that expect a later onset, would be more inclined to purchase the insurance. As expected we find in columns 3 and 4 of Table 7, that the subjective mean of onset is positive and significant at the 5 percent level, when the sample is restricted to households that attended the marketing meeting.[21]

[19] We only have one rainfall gauge with available data in 2004 in Anantapur and 3 in Mahbubnagar. Based again on minimum distance, we reassigned villages to the next closest gauge with available data.

[20] Since weather insurance policies were not advertised outside a marketing meeting, and actual take-up of the product was very low, most farmers that did not attend the meeting would not have heard about weather insurance.

[21] Notice that risk aversion is significant but negative, suggesting that it is *negatively* correlated with the purchase of insurance. This contradictory result is explained in detail in Giné, Townsend and Vickery

Second, we consider in columns 5-8 of Table 7 the average kartis in which the household planted in 2004 and 2006, averaging across crops and plots cultivated. In this case, households that ex-ante believe that the monsoon will arrive later should also plant later. The coefficient on the subjective mean are positive and significant, as well as those of the individual definition and the actual onset of the monsoon. These results are significant because they provide direct evidence that priors do matter, and that there is significant updating. The coefficients suggests that an increase of one kartis in the prior expectation would delay planting by roughly 0.4 kartis (approximately one week).

Next we look in columns 9 and 10 at the percentage of expenditures before the actual onset of the monsoon in 2006. Again, households that believe that the monsoon would start later will end up making fewer purchases before the monsoon actually sets in, although they would have liked to make such purchases in advance. Again, both coefficient of interest have the expected sign and are significant.

Finally, we look at the decision to replant in the last 10 years, a variable collected in 2006. Here, households that believe that the monsoon will start later should have a lower probability of replanting, because households that plant early are most likely to replant. This is exactly what we find.

In sum, prior expectations are a powerful predictor of behavior, even more so than proxies of risk aversion or discount rates. All in all, these results are remarkable because they not only validate the elicitation method, but also indicate that heterogeneity in prior expectations, more so than heterogeneity in preferences, explain actual behavior.

4.3 Who are the good predictors?

We now turn to the test of Proposition 3 and compare the subjective prior distribution of the onset of the monsoon with the historical one, computed using both definitions for the onset of the monsoon. According to Proposition 3, accuracy should depend on how

(2008).

relevant the production of rain-sensitive crops. The specification we run is the following:

$$\log(\chi^2_{ivs}) = Z'_{vs}\alpha + X'_{ivs}\beta + \epsilon_{ivs}$$

where χ^2_{ivs} is the chi-square statistic computed according to the formula in (1) for individual i in village v mapped to rainfall station s. The vector Z_{vs} contains village-level variables such as a district dummy, the proxy for uncertainty V_s described in the previous subsection and the distance from village v to rainfall gauge s. The vector X_{ivs} includes a set of household level characteristics that may affect accuracy. These variables are described in Section 2.4, and include basic demographics, such as caste, literacy and education, wealth, ability to cope with income shocks and exposure to weather shocks. A positive and significant coefficient on the variable X_{ivs} would indicate that X_{ivs} reduces accuracy, since higher χ^2_{ivs} values are associated with larger differences between the prior subjective and historical distribution.

Table 8 reports the results. In columns 1 and 2, the historical distribution is computed using the district average of the individual definitions of the onset of monsoon while in columns 3 and 4, the individual definition is used. Columns 1 and 3 include the variables in Z_{vs}, while columns 2 and 4 include village dummies. The district dummy is significant in column 1, suggesting that after controlling for household characteristics and other village controls, respondents in Anantapur are more inaccurate. The variable distance is significant and positive in column 3, indicating that the possible lower correlation between the timing of rainfall at the gauge and at the village (basis risk) is partly responsible for the observed lack of accuracy.

Variables that proxy for parameters in the utility function do not seem to influence accuracy. The lowest p-value of an F-test that they are jointly insignificant is 0.35.

Other demographic variables such as literacy, age and caste improve accuracy but they are estimated very imprecisely. Thus, accuracy is not about cognition per se, since literacy (or education) do not seem to influence it significantly. An F-test that all demographic variables are jointly insignificant cannot be rejected at the 10 percent (lowest p-value is

0.4).

In addition, the wealthier the household as measured by the log value of the primary dwelling or the log value of the landholdings, the worse the household is in forecasting the monsoon. An F-test that both wealth variables are jointly insignificant is always rejected at the 5 percent. Other variables that proxy for the ability to cope with shocks in general, such as per capita income, participation in a chit fund or being credit constrained are mostly significant and of the expected sign. The F-test is rejected in 3 out of 4 specifications. In addition, variables that are correlated with the exposure to weather shocks, such as the amount of land the household cultivates (positive correlation) and the percentage of land devoted to paddy (negative correlation) are also of the right sign and mostly significant. The F-test is rejected in 2 out of the 4 specifications.

In sum, confirming the prediction of the model, better forecasters seem to be households whose incomes depend more heavily on the monsoon and households that lack proper risk-coping mechanisms to smooth weather shocks.

Thus, households for which the monsoon is important are willing to devote resources (maybe attention?) to acquire more information and thus are on average more accurate at predicting it.

4.4 Does Accuracy Matter?

Finally, we test Proposition 4 in the model. Inaccuracy increases the scope for mistakes in the optimal time to plant, so over time, inaccurate farmers should (i) experience lower yields in rain sensitive crops and (ii) replant fewer times.

We test this proposition by running two regressions. First, the following yield regression

$$y_{cidt} = D_c + D_t + D_i + D_d \times D_c + \alpha_c \log(\chi_i^2) \times D_c + X_c'\beta + \epsilon_{cidt}, \tag{4}$$

where y_{cidt} is the yield per acre of crop c grown by farmer i in district d in year t, D_c is a crop c dummy, D_t is a year dummy, D_i is an individual dummy, $\log(\chi_i^2)$ is the log of the accuracy proxy χ^2 according to the formula in (1), and X_c is a set of crop characteristics, including soil characteristics and use of intercropping, among others. We use crop yields for 2004 and 2006, the years for which we have data.

Second, we run the same specification as that of Table 7 including $\log(\chi_i^2)$ as a regressor.

$$R_i = \alpha_\chi \log(\chi_i^2) + \alpha_\mu \mu_{Si} + \beta r_i + \delta k_i + X_i' \gamma + \epsilon_i,$$

where R_i is the decision to replant in the last 10 years and $\alpha_\chi$is the variable of interest.

Table 9 reports the main results. In Panel A the results from regression (4) with rice as the omitted crop category. Indeed, for all crops, lower accuracy results in lower yields. Because the variable $\log(\chi_i^2)$ is hard to interpret, we conduct the following thought experiment: By how much would income increase for an individual that cultivates rain-sensitive crops using the average land devoted to them, if accuracy was increased from the 25th percentile of the distribution to the 75th percentile? The gains in yields per acre are converted into Kg for each crop using the average land size devoted to them, then Kg are multiplied by the respective prices and they are finally aggregated. We end up with an average gain between Rs 1,500 and Rs 1,800 depending on whether the individual or district definition is used. These income gains translate into 7.8 percent and 9.1 percent of the total value of agricultural production or 5.3 percent and 6.1 percent of total household income.

Panel B shows that accuracy is correlated with a lower probability of having replanted in the past, although the coefficient is only significant when the individual definition of start of monsoon is used to compute the χ^2 measure and actual onset of monsoon. If again accuracy was to increase from the 25th percentile of the distribution to the 75th percentile, the probability of replanting would shrink from XX to YY.

5 Conclusions

Due to the vagaries of the monsoon, rain-fed farmers in semi-arid India should benefit greatly from having accurate expectations about its onset, which determines the optimal time to plant. We collected unique survey data on farmer's subjective probability distribution of the monsoon onset to assess the role that expectations play in decision-making. We then combined these data with historical rainfall data and computed individual measures of accuracy to assess which were the most accurate and why. Finally we quantified the welfare costs of inaccurate expectations.

We show that the elicitation method used delivers informative prior subjective expectations and that there is significant heterogeneity in beliefs across households. More importantly, real outcomes, such as the timing of planting decisions, and agricultural profitability, are strongly influenced by households' beliefs about the monsoon. This suggeststhat expectations are a powerful predictor of behavior, even more so than proxies for parameters in the utility function, such as risk aversion and discount factors.

In addition, we study the determinants of accuracy in expectations and find that individuals with stronger incentives to gather information have more accurate beliefs. As a result, accuracy seems to be less explained by cognitive biases (Kahneman, Slovic and Tversky, 1982; Rabin, 1998) than by how relevant the event is to the forecaster. suggesting that agents use a simple cost-benefit analysis model to acquire and process information.

Finally, we show that accuracy leads to average income gains of 8 to 9 percent of agricultural production, and to a lower probability of replanting, both suggesting that mistakes are costly.

We conclude therefore that eliciting expectations data and incorporating them in decision-making problems seems to be warranted.

References

[1] H. P. Binswanger. Attitudes towards risk: experimental measurement in rural India. *American Journal of Agricultural Economics*, 62(8):395–407, 1980.

[2] H. P. Binswanger. Attitudes towards risk: theoretical implications of an experiment in rural India. *The Economic Journal*, 91(9):867–889, 1981.

[3] H.P Binswanger and D. A. Sillers. Risk Aversion and Credit Constraints in Farmers' Decision Making: A Reinterpretation. *Journal of Development Studies*, 20:5–21, 1984.

[4] A. Delavande and H.P. Kohler. Subjective Expectations in the Context of HIV/AIDS in Malawi. Working Paper, RAND Corporation, 2007.

[5] S. DellaVigna. Psychology and Economics: Evidence from the Field. Working Paper, UC Berkeley, 2007.

[6] W. DuMouchel and G. Duncan. Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association*, 78 (383):535–543, 1983.

[7] J. Fein and P. Stephens. *Monsoons*. John Wiley and Sons, New York, NY, 1987.

[8] S. Gadgil and S. Gadgil. The Indian Monsoon, GDP and Agriculture. *Economic and Political Weekly*, November 25:4887–4895, 2006.

[9] S. Gadgil, P.R. Seshagiri Rao, and K. Narahari Rao. Use of Climate Information for farm-level decision making: rainfed groundnut in southern India. *Agricultural Systems*, 74:431–457, 2002.

[10] X. Giné, R. Townsend, and J. Vickery. Patterns of Rainfall Insurance Participation in Rural India. *The World Bank Economic Review*, 22(3):539–566, 2008.

[11] C. Gollier. *The Economics of Risk and Time.* MIT Press, Cambridge, MA, 2001.

[12] B.N. Goswami, V. Venugopal, D. Sengupta, M.S. Madhusoodanan, and Prince K. Xavier. Increasing Trend of Extreme Rain Events Over India in a Warming Environment. *Science*, 314(5804):1442–1445, 2003.

[13] J. Hirshleifer and J. Riley. *The Analytics of Uncertainty and Information.* Cambridge University Press, Cambridge, UK, 1995.

[14] D. P. Slovic Kahneman and A. Tversky. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, UK, 1982.

[15] S. Klonner. Private Information and Altruism in Bidding Roscas. *The Economic Journal*, 118(528):775–800, 2008.

[16] W. Luseno, J. McPeak, C. Barrett, G. Gebru, and P. Little. The Value of Climate Forecast Information for Pastoralists: Evidence from Southern Ethiopia and Northern Kenya. *World Development*, 31(9):1477–1494, 2003.

[17] T. Lybbert, C. Barrett, J. McPeak, and Winnie Luseno. Bayesian herders: asymmetric updating of rainfall beliefs in response to external forecasts. Working Paper, Cornell University, 2005.

[18] R. Mall, R. Singh, A. Gupta, G. Srinivasan, and L.S. Rathore. Impact of Climate Change on Indian Agriculture: A Review. *Climatic Change*, 78(2-4):445–478, 2006.

[19] C. Manski. Measuring Expectations. *Econometrica*, 72(5):1329–76, 2004.

[20] D. McKenzie, J. Gibson, and S. Stillman. A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? Working Paper, World Bank, 2007.

[21] A. Murphy and R. Winkler. Probability Forecasting in Meteorology. *Journal of the American Statistical Association*, 79(387):489–500, 1984.

[22] K. Narahari Rao, S. Gadgil, P.R. Seshagiri Rao, and K. Savithri. Tailoring strategies to rainfall variability – The choice of the sowing window. *Current Science*, 78(10): 1216–1230, 2000.

[23] P. Norris and R. Kramer. The Elicitation of Subjective Probabilities with Applications in Agricultural Economics. *Review of Marketing and Agricultural Economics*, 58(2-3):127–147, 1990.

[24] W. H. Press. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 1992.

[25] M. Rabin. Psychology and Economics. *Journal of Economic Literature*, 36:11–46, 1998.

[26] M. Rosenzweig and H.P Binswanger. Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *The Economic Journal*, 103(1):56–78, 1993.

[27] P. Singh, K.J. Boote, A. Yogeswara Rao, M.R. Iruthayaraj, A.M. Sheik, S.S Hundal, R.S. Narang, and Phool Singh. . *Field Crops Research*, 39:147–162, 1994.

[28] T Walker and J. Ryan. *Village and Household Economics in Indias Semi-Arid Tropics*. John Hopkins University Press, Baltimore, Maryland, 1990.

Figure 1. District Location in Andhra Pradesh, India

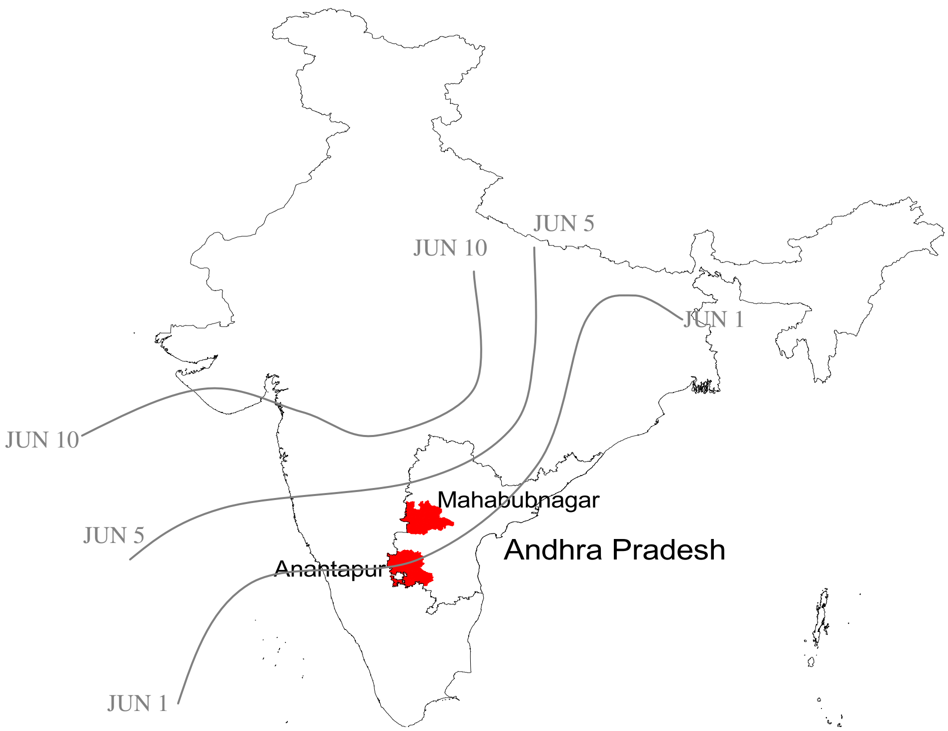


Figure 2. Village Location in Mahbubnagar

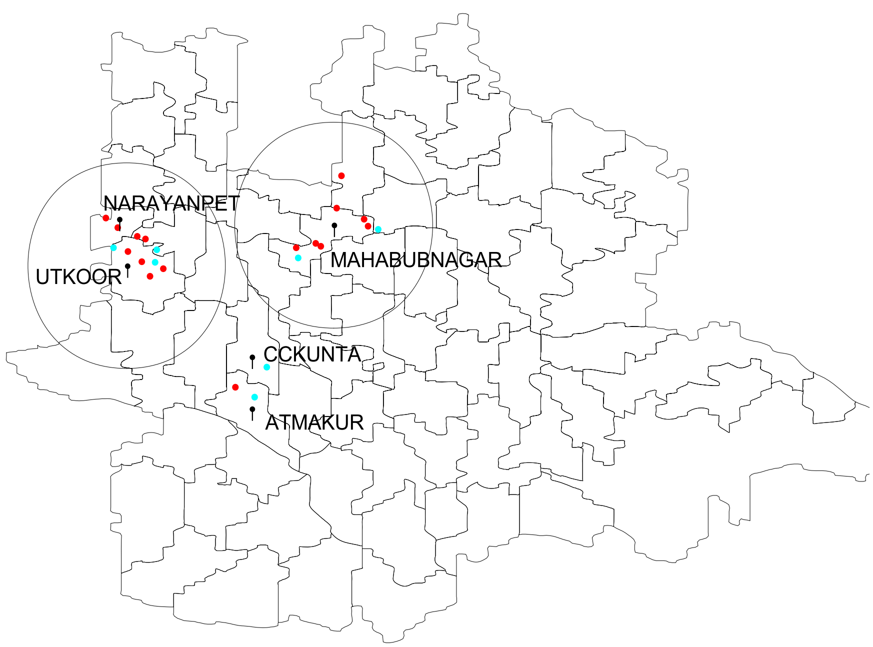


Figure 3. Onset of the Monsoon over Time

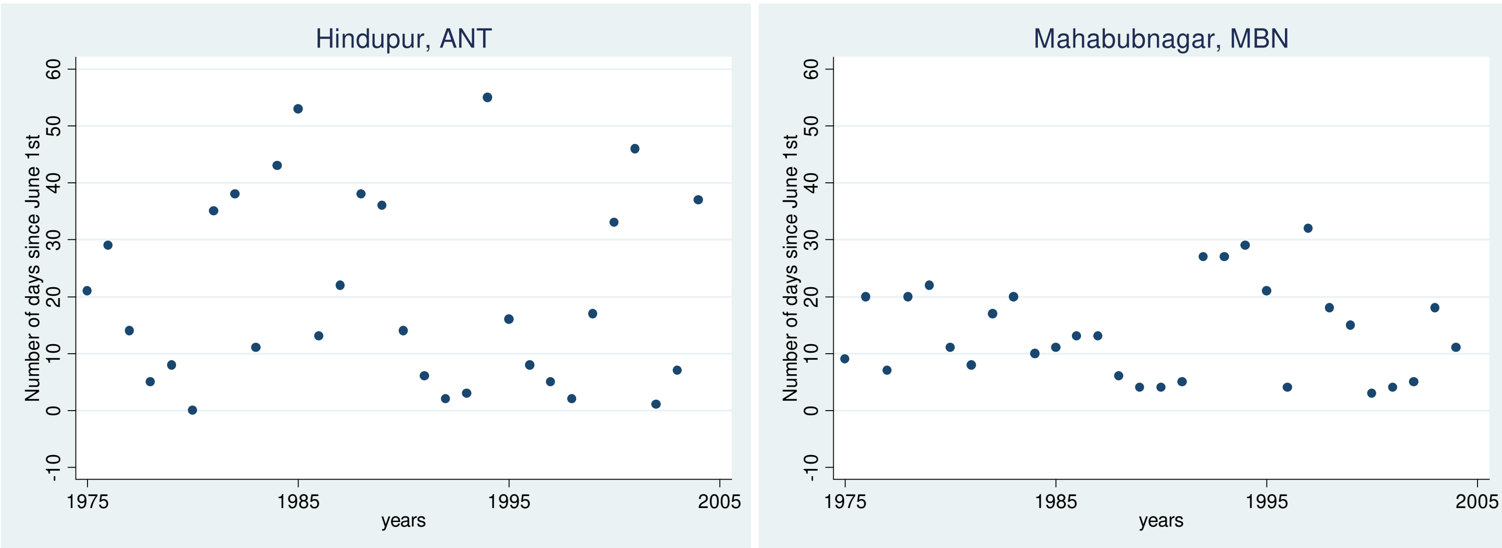


Figure 4. Days until Onset of the Monsoon since IMD Forecast over Time

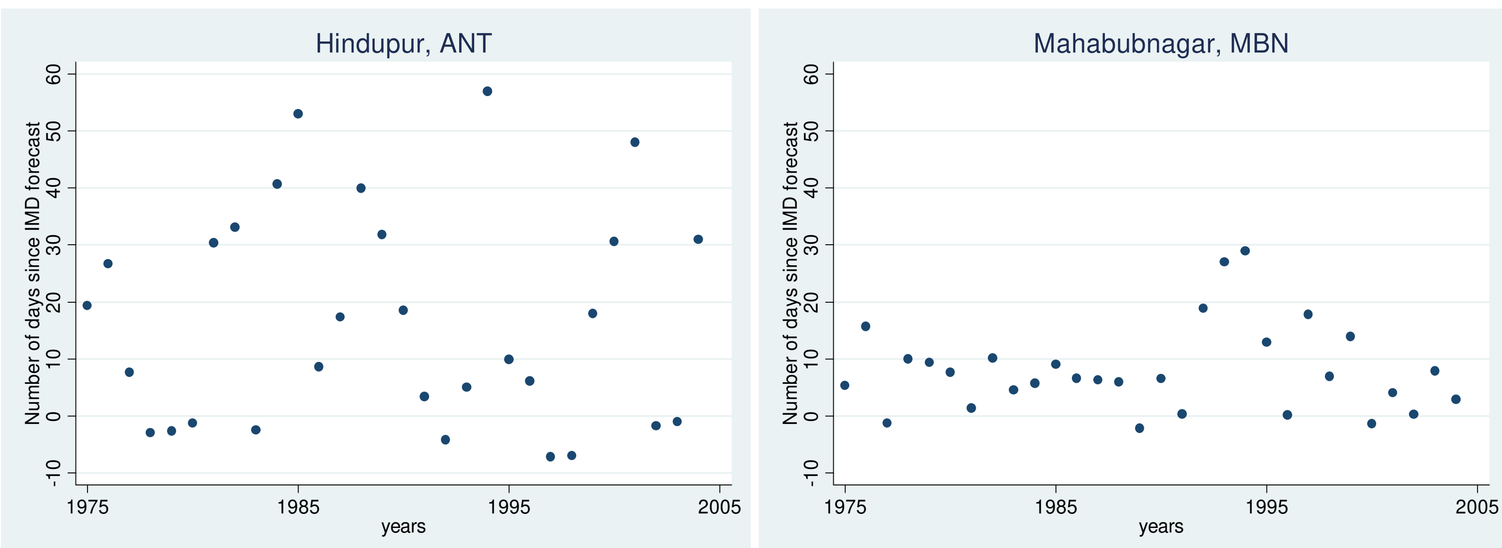


Figure 5. Cumulative Distribution Functions

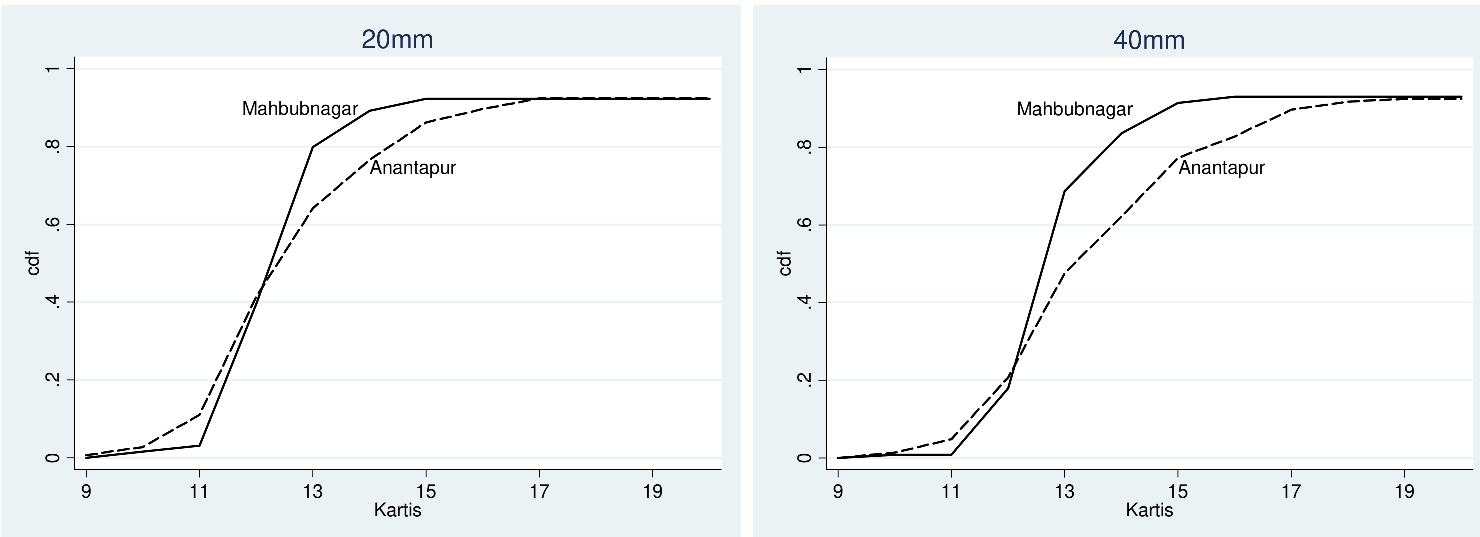


Figure 6. Optimal Planting Strategy

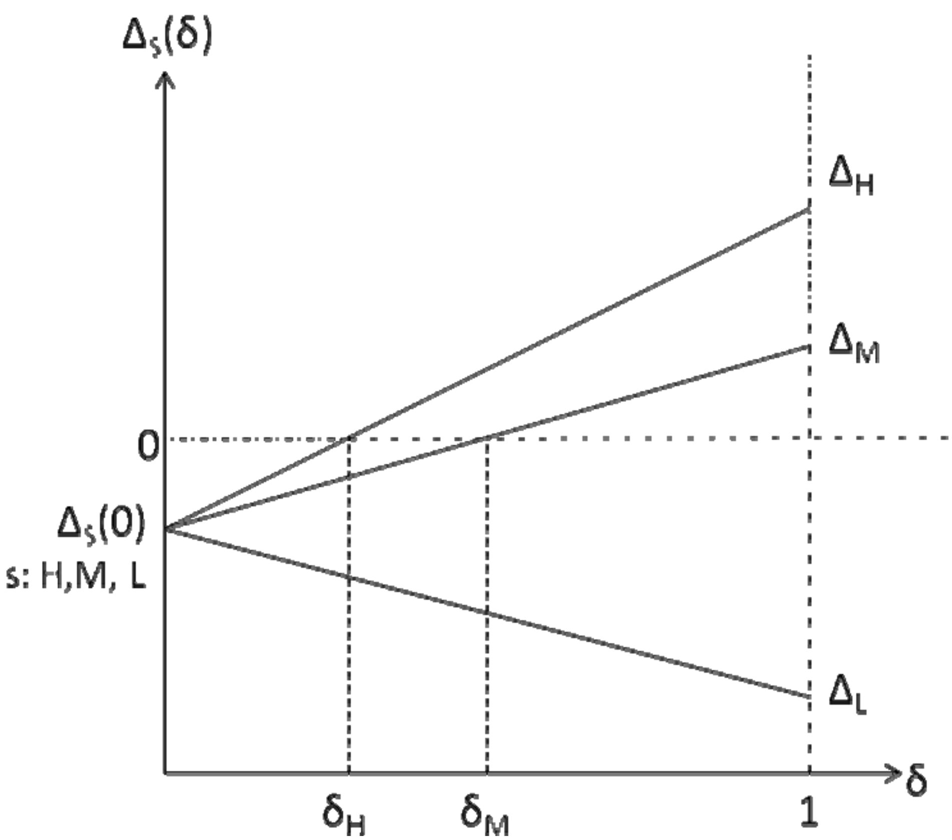


Figure 7: Expected Net Income from Rain-Sensitive Crop

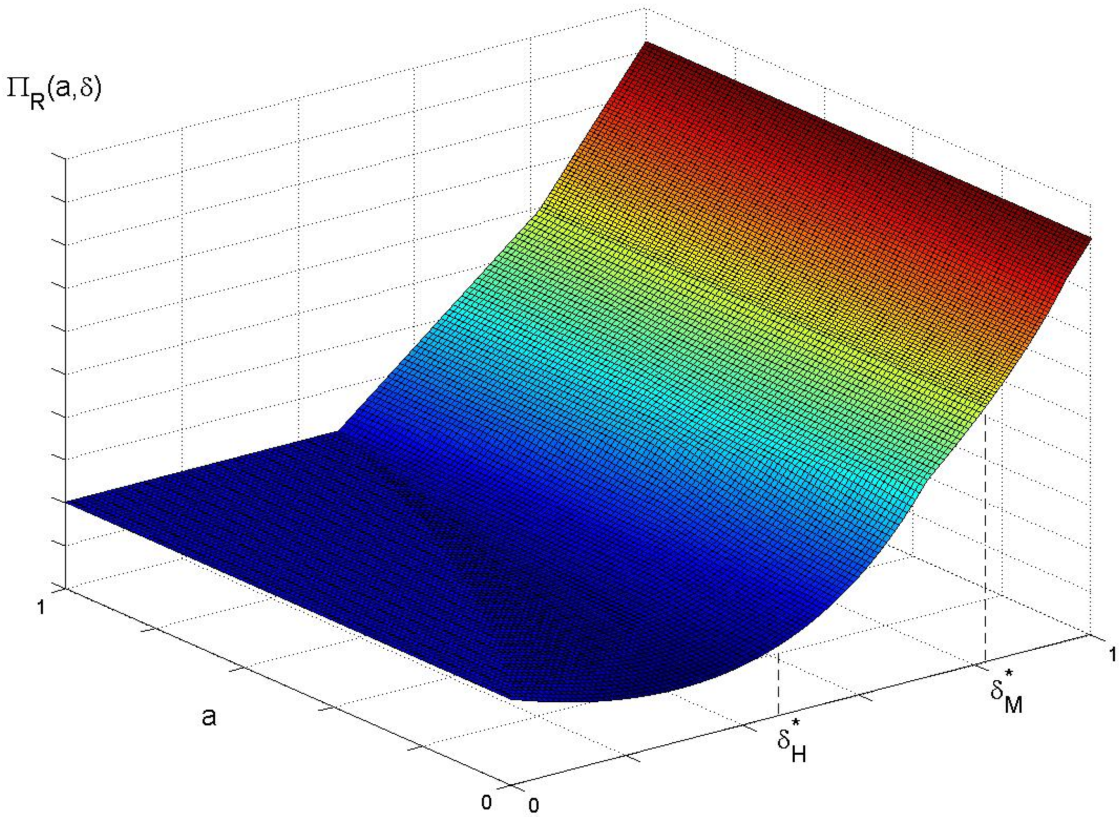
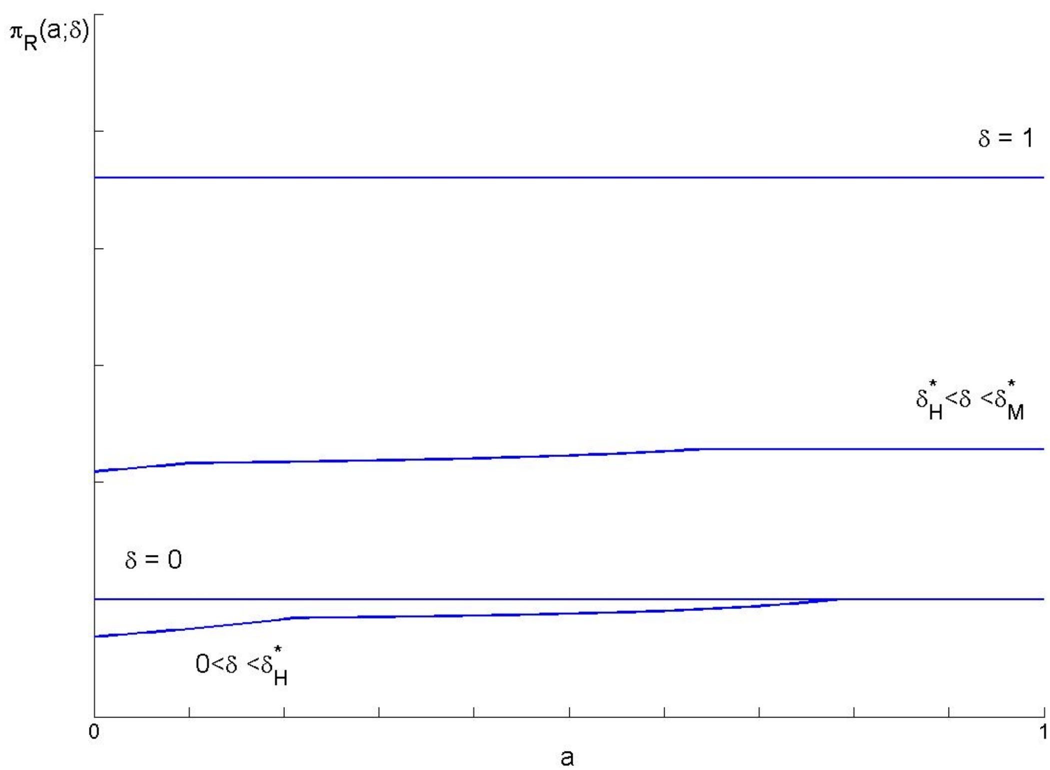


Figure 8: Expected Net Income from Rain-Sensitive Crop as function of *a*



Appendix. Construction of Variables

Variable name	Survey year	Significance of variable
Risk aversion	2004	Dummy variable equal to 1 if respondent selects the safe bet: 50 Rs. if the coin lands on heads and 50 Rs. if the coin lands on tails.
Discount Rate	2004	(100/x-1)*100 where x is the minimum amount they are willing to accept to get a hypothetical lottery price of Rs.1000 immediately instead of waiting for one month.
Membership in BUA (1=Yes)	2004	Dummy variable equal to 1 if any household member belongs to Borewell User Association (BUA) or Water User Group (WUG)
Weather info from informal sources (1=Yes)	2004	Dummy variable equal to 1 if the household actively seeks information about weather from trader/middleman, friend/relative, progressive farmer or neighboring farmer
Market value of the house (Rs.100,000)	2004	Present market value of the household's primary dwelling and the plot located in the house (what would they be able to get if they sell it)
Value of Owned land (Rs. 1.000,000)	2004	Present market value of the owned plots
Participation in chit fund (1=Yes)	2004	Dummy variable equal to 1 if household participates in at least one chit fund (ROSCA)
Household is credit constrained (1=Yes)	2004	Dummy variable equal to 1 if household was denied credit or cites lack of collateral as a reason for not applying or not having one more loan
Per capita total income (Rs. 10,000)	2004	Total income per household member
Cultivated land in acres	2004	Total land cultivated by the household
Pct. of land with paddy	2004	Land dedicated to paddy over total household cultivated land
Pct. land with expenditure in irrigation	2004	Pct of total cultivated acres in which the household have spend on irrigation.
Literacy (1=Yes)	2004	Dummy variable equal to 1 if the household head is literate
Education	2004	Years of education of household's head
Age of the household's head	2004	Age of household's head
Forward caste (1=Yes)	2004	Dummy variable equal to 1 if the household is of forward caste
Pct. of land with loamy sand soil	2004	Pct of total household cultivated land that has loamy sand soil (loamy sand includes alluvial, sandy and red sandy soils)
Pct. of land with loam soil	2004	Pct of total acres cultivated by the HH that have loam soil (red loam soil)
Pct. of land with clay loam	2004	Pct of total acres cultivated by the HH that have clay loam soil (clay loam includes black soil both very shallow and shallow as well as saline soil)
Pct.of land with slope (higher than 1%)	2004	Pct of total acres cultivated by the household that have slope greater than 1%
Pct. of land with soil deeper than 80 cm	2004	Pct of total acres cultivated by the household that are deeper than 80cm.
Individual definition of start of monsoon (mms)	2004	Amount of rain (mms) required to start sowing. The most of the respondents report the amount of soil moisture that they need and we used the following conversion to get the amount in mms: 4.318 mms./in. for loamy sand soil, 6.858 mms./in. for loam, 5.842 mms./in. for silty clay soil and 8.382 mms./in. for clay loam soil.
Pct. of expenditures before onset of the monsoon	2006	Percentage of total inputs' expenditure invested before the onset of the monsoon in 2006
Dummy Anantapur	2004	1 if household is located in the district of Anantapur
Standard deviation of onset of monsoon	-	Standard deviation of number of days since June 1st until the start of the monsoon for each station across time.
Distance from village to rainfall gauge	2004	Distance from village to rainfall gauge in kms.
Age of the eldest household member > 60 (1=Yes)	2004	1 if the eldest member of the household is more than 60 years old.
Attended meeting and bought insurance in 2004	2004	Dummy variable equal to 1 if the household attended the weather insurance meeting and bought insurance in 2004
Bought insurance in 2004	2004	Dummy variable equal to 1 if the household bought rainfall insurance for 2004
Average planting kartis	2004 & 2006	Average kartis when planting took place across crops in relevant year
HH replanted in last 10 years (1=Yes)	2006	Dummy variable equal to 1 if the household has ever replanted at least one crop
Subjective mean of onset	2004	Mean kartis of the subjective distribution
Actual onset of monsoon	2004 & 2006	Actual onset of the monsoon in relevant year computed using either the individual definition or the average of the individual definitions in the district.

Table 1. Assignment of villages to rainfall stations in each district

Panel A: Mahbubnagar				
Rainfall gauge	N. Villages	N. Households	Distance (miles)	N. Years
Utkor	6	204	4.01	17
Narayanpet	5	142	2.87	25
Mahbubnagar	9	247	5.33	35
Atmakur	1	25	2.53	35
CC Kunta	2	70	3.44	17
Total	23	688	4.14	25.8
Panel B: Anantapur				
Rainfall gauge	N. Villages	N. Households	Distance (miles)	N. Years
Somandepalli	4	111	5.22	14
Madakasira	3	70	5.63	34
Hindupur	1	26	0.69	35
Parigi	3	80	3.58	16
Lepakshi	3	76	4.74	16
Total	14	363	4.51	20.2

Notes: Each village is assigned to the closest rainfall gauge in the district. "N. Villages" is the number of villages assigned to the rainfall gauge. "N. Households" is the number of households interviewed in the villages assigned to the rainfall gauge. "Distance in miles" is the average distance from the rainfall gauge to the villages assigned to the rainfall gauge. "N. Years" is the average number of years with available data. The row "Total" reports the district averages of "Distance" and "N. Years" weighted by number of households assigned to each rainfall gauge, and district totals of "N. Villages" and "N. Households".

Table 2. Evidence of inverse U-shape between Yield and Planting Time

			Anantapur		
	(1)	(2)	(3)	(4)	(5)
	Paddy	Groundnut	Castor	Redgram	Sorghum
Kartis	-4050.477	-1885.659	513.019	-2045.68	290.972
	[1998.007]**	[2966.379]	[144.242]***	[2944.891]	[114.674]**
$Kartis^2$	127.975	57.827	-25.229	63.076	-9.517
	[68.041]*	[96.880]	[6.340]***	[96.066]	[3.285]***
			Mahbubnagar		
	(1)	(2)	(3)	(4)	(5)
	Paddy	Groundnut	Castor	Redgram	Sorghum
Kartis	-283.467	541.809	790.719	933.063	-61.235
	[414.824]	[268.121]**	[454.546]*	[382.756]**	[224.044]
$Kartis^2$	9.84	-18.955	-30.09	-34.673	1.441
	[14.285]	[8.250]**	[16.708]*	[14.010]**	[7.855]
Observations			2622		
R-squared			0.89		

Notes: Reported in the different cells are the coefficients of the interaction of crop x Kartis or crop x Kartis squared from a regression of yields per acre. The regression includes crop x district dummies and plot level characteristics such as whether Soil is loamy sand (1=Yes), Soil is loam (1=Yes), Soil is clay loam (1=Yes), Soil is deeper than 80cm. (1=Yes) and Use of Intercropping (1=Yes). Each observation is a subplot. The regression is estimated using OLS with household fixed effects. Robust standard errors clustered at individual level are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent level, respectively.

Table 3. Kartis Codes

Code	Name	Dates
9	Ashwini	Apr 13 – Apr 26
10	Bharani	Apr 27 – May 10
11	Krittika	May 11 – May 23
12	Rohini	May 24 – June 6
13	Mrigashira	June 7 – June 20
14	Ardra	June 21 – July 5
15	Punarvasu	July 6 – July 19
16	Pushya	July 20 – Aug 2
17	Aslesha	Aug 3 – Aug 16
18	Makha	Aug 17 – Aug 29
19	Pubbha	Aug 30 – Sep 12
20	Uttara	Sep 13 – Sep 25

Notes: The code is a serial number that takes value 1 with the first kartis of the year.

Table 4: Comparison of Subjective and Historical Distributions

	All			Mahbubnagar			Anantapur		
	Subjective	Historical		Subjective	Historical		Subjective	Historical	
		Individual	District		Individual	District		Individual	District
	Means								
Support of Distributions									
Highest bin	15.16	16.21	15.57	14.95	15.39	14.85	15.56	17.77	16.95
Lowest bin	12.60	12.69	12.00	12.23	12.80	12.00	13.31	12.47	12.00
Distribution has ... with positive mass (1= Yes)									
1 bin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 bins	3.34	2.14	0.00	0.15	3.29	0.00	9.34	0.00	0.00
3 bins	43.9	30.2	26.3	33.7	46.1	40.2	63.0	0.37	0.00
4 bins	46.5	37.9	31.2	57.7	45.6	37.6	25.5	23.4	19.2
5 bins	5.79	13.9	27.0	8.16	5.07	22.3	1.35	30.6	35.9
6 bins	0.49	8.36	15.6	0.32	0.00	0.00	0.80	24.1	44.8
7 bins	0.00	6.27	0.00	0.00	0.00	0.00	0.00	18.04	0.00
8 bins	0.00	1.20	0.00	0.00	0.00	0.00	0.00	3.46	0.00
9 bins	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.10	
Moments and Properties of Distributions									
Mean	13.8	13.8	13.4	13.5	13.7	13.3	14.4	14.2	13.7
Standard Deviation	0.87	0.97	1.02	0.91	0.76	0.85	0.79	1.37	1.33
Unimodal (1=Yes)	93.1	70.4	77.9	92.1	96.2	93.2	95.1	21.9	49.1
Accuracy									
Subjective dist is different from historical at…									
1 percent		0.13	0.07	-	0.19	0.07	-	0.03	0.08
5 percent		0.26	0.17	-	0.32	0.23	-	0.13	0.13
10 percent		0.33	0.25	-	0.40	0.42	-	0.20	0.16
Subjective dist has same mode as historical (1=Yes)									
		0.43	0.48	-	0.51	0.26	-	0.30	0.59

Notes: Observations are weighted by the appropiate stratification (Attending Marketing Meeting and Purchasing Insurance in 2004) weights.

Table 5. Summary Statistics

	Full Sample				Mahbubnagar	Anantapur	Difference
	Mean	Std. Dev.	Min	Max	Means		
Utility Function							
Risk aversion	0.37	0.48	0.00	1.00	0.38	0.36	-0.015
Discount Rate	29.5	28.4	0.00	233	25.7	36.8	11.060***
Information / Social Networks							
Membership in BUA (1=Yes)	0.03	0.16	0.00	1.00	0.04	0.00	-0.038*
Weather info from informal sources (1=Yes)	0.19	0.39	0.00	1.00	0.18	0.22	0.038
Wealth							
Market value of the house (Rs.100,000)	0.69	0.59	0.06	5.00	0.72	0.64	-0.078
Value of Owned land (Rs. 1.000,000)	0.25	0.42	0.00	21.0	0.31	0.13	-1.722***
Ability to smooth income shocks							
Participation in chit fund (1=Yes)	0.24	0.43	0.00	1.00	0.19	0.33	0.142
Household is credit constrained (1=Yes)	0.80	0.40	0.00	1.00	0.80	0.80	-0.003
Per capita total income (Rs. 10,000)	0.57	2.8	0.02	62.5	0.57	0.57	-0.019
Exposure to Rainfall Shocks							
Cultivated land in acres	6.28	6.03	0.00	82.0	6.93	5.06	-1.871***
Pct. of land with paddy	0.06	0.14	0.00	1.00	0.09	0.01	-0.089***
Pct. land with expenditure in irrigation	0.10	0.20	0.00	1.00	0.09	0.12	0.031
Other Household Characteristics							
Literacy (1=Yes)	0.37	0.48	0.00	1.00	0.33	0.46	0.133**
Education	3.30	4.36	0.00	18.0	2.90	4.04	1.140**
Age of the household's head	47.2	11.4	21.0	80.0	47.8	46.1	-1.756
Forward caste (1=Yes)	0.15	0.36	0.00	1.00	0.12	0.21	0.090
Land Characteristics							
Pct. of land with loamy sand soil	0.50	0.45	0.00	1.00	0.51	0.48	-0.029
Pct. of land with loam soil	0.21	0.38	0.00	1.00	0.16	0.31	0.147*
Pct. of land with clay loam	0.19	0.35	0.00	1.00	0.22	0.13	-0.083
Pct.of land with slope (higher than 1%)	0.45	0.46	0.00	1.00	0.48	0.41	-0.069*
Pct. of land with soil deeper than 80 cm	0.12	0.27	0.00	1.00	0.13	0.09	-0.035
Cropping Patterns of Rainfed Crops [1]							
Number of Plots used	1.86	0.81	1.00	7.00	2.03	1.54	-0.486***
Number of Crops cultivated	2.13	0.79	1.00	5.00	2.16	2.08	-0.071
Farmer plants all crops in one kartis (1=Yes)	0.71	0.45	0.00	1.00	0.57	0.96	0.389***
Planting Kartis across crops	14.06	0.96	12.00	17.00	13.60	14.91	1.309***
Number of Observations	980				650	330	
Number of Observations	1,048				686	362	

Notes: Observations are weighted by the appropriate stratification (Attending Marketing Meeting and Purchasing Insurance in 2004) weights. [1] The variables are computed including the most popular rainfed crops: groundnut, castor, sorghum and redgram. The units in Planting Kartis is the kartis number. See Table 3 for a list of kartis. The Appendix contains a detailed description of the rest of the variables.

Table 6. Irreversible Investments

	Rainfall required to plant (mms.)			Pct. expenses before monsoon		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy Anantapur	5.124		1.864	-0.282		-0.286
	[0.624]***		[1.156]	[0.011]***		[0.019]***
Conditional St. Dev. of monsoon onset		0.436	0.311		-0.019	0
		[0.050]***	[0.094]***		[0.001]***	[0.002]
Risk aversion	0.423	0.442	0.452	0.014	0.016	0.014
	[0.511]	[0.508]	[0.508]	[0.010]	[0.011]	[0.010]
Discount rate (*100)	0.489	1.059	0.802	-0.022	-0.061	-0.021
	[0.939]	[0.906]	[0.936]	[0.019]	[0.019]***	[0.019]
Pct. of land with loamy sand soil	-2.98	-2.645	-2.817	-0.079	-0.105	-0.079
	[2.625]	[2.580]	[2.620]	[0.048]*	[0.052]**	[0.048]
Pct. of land with loam soil	8.528	8.757	8.57	-0.07	-0.098	-0.07
	[2.741]***	[2.690]***	[2.732]***	[0.048]	[0.053]*	[0.048]
Pct. of land with clay loam	16.48	16.008	16.127	-0.065	-0.047	-0.066
	[2.528]***	[2.489]***	[2.506]***	[0.046]	[0.051]	[0.046]
Pct.of land with slope (higher than 1%)	-1.234	-1.327	-1.306	-0.008	-0.005	-0.008
	[0.544]**	[0.546]**	[0.546]**	[0.011]	[0.012]	[0.011]
Pct. of land with soil deeper than 80 cm	9.936	9.591	9.642	-0.048	-0.04	-0.048
	[2.449]***	[2.390]***	[2.417]***	[0.043]	[0.048]	[0.043]
Pct. land with expenditure in irrigation	1.338	0.123	0.29	0.02	0.045	0.019
	[1.596]	[1.677]	[1.681]	[0.028]	[0.029]	[0.029]
Pct. of land with paddy	-3.347	-2.197	-2.024	-0.009	0.019	-0.008
	[2.130]	[2.224]	[2.208]	[0.044]	[0.047]	[0.045]
Forward caste (1=Yes)	1.145	1.828	1.57	-0.017	-0.056	-0.017
	[0.803]	[0.815]**	[0.802]*	[0.014]	[0.014]***	[0.014]
Education	-0.042	-0.02	-0.029	-0.001	-0.003	-0.001
	[0.063]	[0.062]	[0.063]	[0.001]	[0.001]**	[0.001]
Age of the household's head	0.033	0.039	0.037	-0.001	-0.001	-0.001
	[0.021]	[0.021]*	[0.021]*	[0.000]***	[0.000]***	[0.000]**
Cultivated land in acres	0.054	0.032	0.043	-0.001	0.001	-0.001
	[0.046]	[0.047]	[0.047]	[0.001]	[0.001]	[0.001]
Mean dependent variable	30.35	30.35	30.35	0.23	0.23	0.23
Observations	1045	1045	1045	1045	1045	1045
R-squared	0.48	0.48	0.48	0.46	0.36	0.46

Notes: In columns (1)-(3) the dependent variable is the amount of rain in millimeters that respondents require to start planting. In colums (4)-(6) the dependent variable is the percentage of all agricultural expenditures made before the onset of the monsoon. The expenses include bullock and manual labor, hiring tractors, manure, irrigation, purchase of seeds and fertilizer. Standard deviation of Monsoon Onset is the standard deviation of the number of days since 1st June until the onset of monsoon (using individual definition) across years. All regressions control for the variables used in the stratification. Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Table 7. Do people behave according to their expectations?

	Individual (1)	District (2)	Individual (3)	District (4)	Individual (5)	District (6)	Individual (7)	District (8)	Individual (9)	District (10)	Individual (11)	District (12)
Subjective mean of onset	-0.015	-0.015	0.078	0.08	0.39	0.392	0.401	0.402	-0.062	-0.062	-0.082	-0.083
	[0.001]***	[0.001]***	[0.009]***	[0.009]***	[0.056]***	[0.056]***	[0.062]***	[0.063]***	[0.010]***	[0.010]***	[0.021]***	[0.021]***
Ind. def of monsoon (mms.)	0.001	0.001	-0.002	0	0.004	0.007	-0.009	0.009	0	-0.001	0.001	-0.002
	[0.000]***	[0.000]***	[0.001]***	[0.001]	[0.003]	[0.003]**	[0.005]*	[0.004]**	[0.001]	[0.001]**	[0.002]	[0.002]
Actual onset of monsoon	0.001	0.003	0.076	0.057	0.142	0.117	0.928	0.725	-0.057	-0.052	-0.182	-0.138
	[0.001]	[0.001]**	[0.007]***	[0.006]***	[0.029]***	[0.028]***	[0.106]***	[0.090]***	[0.020]***	[0.018]***	[0.056]***	[0.049]***
Risk aversion	-0.031	-0.031	-0.107	-0.104	-0.106	-0.101	-0.102	-0.092	0.022	0.022	0.003	0.003
	[0.002]***	[0.002]***	[0.014]***	[0.015]***	[0.061]*	[0.061]*	[0.099]	[0.100]	[0.014]	[0.014]	[0.038]	[0.038]
Discount Rate (*100)	-0.002	0	-0.116	-0.118	-0.024	-0.027	-0.08	-0.03	-0.053	-0.057	0.073	0.063
	[0.005]	[0.005]	[0.025]***	[0.025]***	[0.117]	[0.117]	[0.161]	[0.162]	[0.026]**	[0.025]**	[0.065]	[0.065]
Pct. of land with paddy	0.269	0.272	-0.277	-0.291	-0.496	-0.502	-0.791	-0.857	0.151	0.152	0.078	0.092
	[0.010]***	[0.011]***	[0.033]***	[0.033]***	[0.184]***	[0.187]***	[0.233]***	[0.236]***	[0.044]***	[0.043]***	[0.117]	[0.118]
Credit constrained (1=Yes)	-0.018	-0.018	-0.006	0.003	-0.027	-0.023	-0.023	-0.043	0	0.002	0.095	0.099
	[0.003]***	[0.003]***	[0.017]	[0.017]	[0.079]	[0.079]	[0.121]	[0.121]	[0.015]	[0.015]	[0.039]**	[0.039]**
Mean dependent variable	0.56	0.56	0.45	0.45	14.18	14.18	14.21	14.21	0.25	0.25	0.30	0.30
Observations	57333	57333	5043	5043	1033	1033	917	917	739	739	666	666
R-squared	0.04	0.04	0.13	0.12	0.17	0.17	0.18	0.17	0.28	0.28	0.13	0.13

Notes: "Actual Start of Monsoon" in odd-numbered columns is the kartis in which accumulated rainfall for the relevant year when the dependent variable was measured reached the use the "individual" definition. In even-numbered columns, the "district" average of individual definitions, rather than each individual definition is used. In columns (1) and (2), regressions are weighted. In columns (3)-(12), regressions control for variables used in the stratification (attended marketing meeting and insurance purchase). In addition, all regressions include the following controls: Forward caste (1=Yes), Literacy (1=Yes), Age of the household's head, Market value of the house (Rs.100,000), Value of Owned land (Rs. 1,000,000), Use of intercropping system (1=Yes). Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Table 8. Differences between subjective and historical distributions

	District average		Individual average	
	(1)	(2)	(4)	(5)
Village characteristics				
Dummy Anantapur	0.999		-0.049	
	[0.106]***		[0.069]	
Conditional St. Dev. of monsoon onset	-0.036		0.01	
	[0.009]***		[0.006]*	
Distance from village to rainfall gauge	0.006		0.065	
	[0.020]		[0.013]***	
Utility Function				
Risk aversion	-0.054	-0.081	0.05	0.032
	[0.063]	[0.061]	[0.047]	[0.045]
Discount Rate	0.001	0.001	0.001	0.001
	[0.001]	[0.001]	[0.001]	[0.001]
Wealth				
Logarithm of market value of the house (Rs.100,000)	0.109	0.091	0.083	0.084
	[0.045]**	[0.044]**	[0.036]**	[0.036]**
Logarithm of value of Owned land (Rs. 1.000,000)	0.11	0.034	0.142	0.11
	[0.058]*	[0.060]	[0.040]***	[0.044]**
Ability to smooth income shocks				
Participation in chit fund (1=Yes)	0.198	0.081	0.234	0.152
	[0.067]***	[0.074]	[0.053]***	[0.058]***
Household is credit constrained (1=Yes)	-0.105	-0.08	-0.136	-0.134
	[0.067]	[0.064]	[0.050]***	[0.047]***
Logarithm of per capita total income (Rs. 10,000)	0.078	0.029	0.042	0.016
	[0.042]*	[0.041]	[0.038]	[0.037]
Exposure to Rainfall Shocks				
Logarithm of cultivated land in acres	-0.166	-0.008	-0.254	-0.158
	[0.067]**	[0.073]	[0.054]***	[0.059]***
Pct. of land with paddy	0.115	0.3	0.103	0.082
	[0.242]	[0.248]	[0.171]	[0.170]
Information / Social Networks				
Membership in BUA (1=Yes)	-0.081	0.277	-0.387	-0.065
	[0.115]	[0.154]*	[0.091]***	[0.168]
Weather info from informal sources (1=Yes)	-0.042	-0.003	-0.15	-0.108
	[0.079]	[0.078]	[0.068]**	[0.064]*
Other Household Characteristics				
Literacy (1=Yes)	-0.083	-0.047	0.006	0.038
	[0.069]	[0.066]	[0.050]	[0.047]
Age of the household's head	-0.003	-0.002	0.022	0.01
	[0.003]	[0.003]	[0.022]	[0.022]
Age of the eldest household member > 60 (1=Yes)	-0.08	-0.07	-0.019	-0.032
	[0.067]	[0.064]	[0.050]	[0.048]
Forward caste (1=Yes)	-0.068	-0.078	0.022	0.063
	[0.081]	[0.081]	[0.059]	[0.059]
Village Fixed Effects	No	Yes	No	Yes
Mean dependent variable	1.50	1.50	1.99	1.99
Observations	1031	1031	1031	1031
R-squared	0.19	0.31	0.13	0.26

Notes: Dependent variable is the logarithm of the chi-square computed from comparing the subjective and the historical distributions. In colums (1) and (2) the historical distribution is computed using the average in the district of the minimum quantum of rainfall that respondent requires to plant while columns (3) and (4) use individual definitions of the onset of the monsoon. All regressions control for the variables used in the stratification. Columns (2) and (4) include village fixed effects. Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Table 9. Does Accuracy Matter?

	Individual	District
	(1)	(2)
Panel A: Yields		
Log ($\chi^{2)}$ * Sorghum	-80,480	-60.962
	[29.316]***	[21.876]***
Log (χ2) * Groundnut	-100,772	-65.355
	[26.672]***	[23.382]***
Log (χ2) * Castor	-84,103	-65.198
	[26.552]***	[20.349]***
Log (χ2) * Redgram	-84,122	-63.855
	[25.998]***	[21.341]***
Number of Observations	3726	3726
R squared	0.67	0.67
Panel B: Household ever Replanted		
Log(χ^2)	0.051	0.029
	[0.023]**	[0.021]
Subjective mean of onset	-0.081	-0.1
	[0.020]***	[0.025]***
Ind. def of monsoon (mms.)	0	-0.003
	[0.002]	[0.002]
Actual onset of monsoon	-0.14	-0.04
	[0.060]**	[0.019]**
Observations	666	666
R-squared	0.14	0.13

Notes: In column (1) the Chi squared is computed using the historical distribution using the average in the district of the minimum quantum of rainfall that respondent requires to plant while column (2) use individual definitions of the onset of the monsoon. In Panel A, the regression includes crop dummies, crop x district dummies and plot level characteristics such as whether Soil is loamy sand (1=Yes), Soil is loam (1=Yes), Soil is clay loam (1=Yes), Soil is deeper than 80cm. (1=Yes) and Use of Intercropping (1=Yes). Each observation is a crop. The excluded category is rice. The regression is estimated using OLS with household fixed effects and clustering at household level. In Panel B, "Actual Start of Monsoon" in column (1) is the kartis in which accumulated rainfall for the relevant year when the dependent variable was measured reached the use the "individual" definition. In column (2), the "district" average of individual definitions, rather than each individual definition is used. The regressions control for variables used in the stratification (attended marketing meeting and insurance purchase). In addition, the following controls are included: Risk aversion, Discount Rate, Pct of Land with Paddy, Credit Constraints (1=Yes), Forward caste (1=Yes), Literacy (1=Yes), Age of the household's head, Market value of the house (Rs.100,000), Value of Owned land (Rs. 1,000,000) and Use of intercropping system (1=Yes). For both panels, robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent level, respectively.