Rational Expectations?
Evidence from Planting Decisions in Semi-Arid India

Xavier Giné* — The World Bank
Robert M. Townsend — U. of Chicago
James Vickery — New York Fed

First draft: June 11, 2007
This draft: October 22, 2007

Abstract

In regions of semi-arid India where cultivation is mostly rainfed, the optimal time to plant is at the onset of the monsoon. While planting too early may result in losses because the seeds may not germinate if the first rains are not followed by subsequent rains, postponing planting for too long is also costly, because yield will typically be lower. Because mistakes about the precise timing of the monsoon appear

*Financial support from CRMG, World Bank and GARP 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 are also grateful to Don Larson for his encouragement, and to Adelaine Delavande, Lance Lochner, David Mckenzie and participants in several conferences and seminars for valuable comments. Paola de Baldomero Zazo, Helene Bie Lilleor 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@uchicago.edu`, and `james.vickery@ny.frb.org`.

to be costly, in this paper we study the differences in the accuracy of subjective expectations that farmers have regarding the timing of the onset of the monsoon. We find that (i) in the region where rains are more erratic, the amount of rainfall that triggers planting is larger, consistent with the predictions of a real options model with differing signal quality; (ii) farmers that have less access to weather risk coping mechanisms forecast the onset more accurately, consistent with agents form expectations by acquiring information via a simple cost-benefit analysis; and finally, (iii) farmers make decisions in accordance to their expectations.

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, risk would not be a source of concern because 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 way to reduce the exposure to weather risk, particularly if agricultural production is rainfed, is to choose an optimal sowing window (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, although they may decide to abandon the crop altogether. Either strategy results in significant losses. But 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 the 1,054 farmers in our sample 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 predictions of the onset of the monsoon.[1]

In this paper we use unique survey data to assess how well farmers predict the start of the monsoon. In particular, we study whether farmers have the same beliefs, and if not, 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 definition was obtained by giving each respondent 10 stones and a sheet of paper with boxes corresponding to 13-14 day periods called "kartis" based on the traditional solar calendar (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. Using the historical rainfall data from the nearby rainfall gauge, we compute for each respondent a chi-square statistic that compares the elicited subjective distribution to the corresponding historical one constructed using the respondent's own definition of

[1] Although farmers could plant different plots at different times as a risk-coping strategy, we show that most farmers plant main rain sensitive crops at the same time.

the onset of the monsoon.

We show that this elicitation method delivers informative subjective expectations and that individuals behave according to them. First, in the district where rains are more erratic, farmers require a significantly larger quantum of rainfall (or depth of soil moisture) to start planting, which suggests that farmers behave rationally and according to the predictions of a real options model with differing signal quality. 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) than in the case when the first rains are uninformative. Second, farmers seem to form expectations by acquiring information as 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 more heavily dependant on it. Thus, accuracy is more explained by how relevant the event is to the forecaster, than by heuristics or "rules of thumb" developed in the psychology and behavioral economics literature (Kahneman, Slovic and Tversky, 1982; Rabin, 1998 for a review).

Finally, we show that farmers who believe the monsoon will start later are also more likely to plant later, 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 expectations, even when controlling for self-reported proxies of risk aversion and discount rates. 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.

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). 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 do not have prior 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.

The rest of the paper is organized as follows. The next section describes the context and survey and weather data collected. Section 3 presents the empirical strategy and the results and finally Section 4 concludes.

2 Data and Context

This paper uses two sources of data, survey data and historical rainfall data. The survey data was collected for the baseline of an impact evaluation of the weather insurance pilot described in Giné, Townsend and Vickery (2007). 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 the policy in 2004. All summary statistics report weighted data by the appropriate population weights (again see Giné, Townsend and Vickery, 2007 for further details). Throughout the analysis, if "purchase of insurance" is not the dependent variable, the stratification variables (attended meeting and purchasing insurance) will always be included as controls. When "purchased insurance" is used as the dependent variable, we will follow Manski and Lerman (1977) and use a weighted regression.[2]

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 hopefully 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.[3] Farmers grow a variety of cash and subsistence crops that vary in their yield's sensitivity to

[2]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. See DuMouchel and Duncan (1983) for details.

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

drought. The main cash crops grown in the area are castor which is mostly grown in Mahbubnagar and covers 34 percent of its cultivable land in the sample, and groundnut, mostly grown in Anantapur and covering half of the cultivable land in the Anantapur sample, paddy (close to 9 percent of the combined cultivated land), maize (4.6 percent), sunflower (4 percent) and cotton (3 percent). Paddy is almost exclusively irrigated (84 percent of all plots use irrigation) and to a lesser extent, maize, sunflower and cotton are also irrigated (roughly one fifth of the plots use irrigation). The subsistence crops grown in the area are grams (redgram and to a lesser extent greengram, covering 10.6 percent of all cultivated land) sorghum (6.7 percent) and millet (1.45 percent). All subsistence crops as well as castor and groundnut are typically rainfed (less than 5 percent of plots are irrigated). Paddy is exclusively grown as a sole crop, sorghum is grown alone in 70 percent of the 204 plots devoted to it in Mahbubnagar and roughly 85 percent of the 45 plots in Anantapur. Groundnut is mostly grown alone in Mahabubnagar (94 percent of 81 plots), but mostly intercropped with redgram in Anantapur (94.5 of 422 plots). Castor is also grown intercropped with redgram in 40 percent of the 605 plots in Mahbubnagar. Finally redgram is typically grown intercropped with either groundnut or castor. Only in 27 percent of 453 crops is redgram cultivated alone in Mahbubnagar, while virtually all the 439 plots with redgram in Anantapur are intercropped.[4]

When irrigation cannot be used to smooth the vagaries of the monsoon, farmers try to choose an optimal sowing window (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. In the former case, farmers should postpone planting because if they were to plant, the seeds would not

[4]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.

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 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 the two main subsistence crops (sorghum and redgram) against the kartis and the kartis squared when the crop was planted. We use kartis instead of weeks or months because it is the unit of time that farmers typically use (Rao et al., 2000; Gadgil et al., 2002). Table 3 reports the relevant kartis for the main cropping season measured by a serial number that takes value 1 for the first kartis of the year. Columns 1 and 3 of Table 2 do not include controls, while columns 2 and 4 control for plot level characteristics. As discussed above, in all specifications but the last one, the linear and quadratic terms are 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.

Before the monsoon arrives, 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. Table 4 reports the average amount per household of all production expenditures that were incurred before the end of April (Bharani) and before the start of the monsoon in 2006. As we can see, they amount to a sizeable one third of all production costs in Mahabubnagar district but virtually none in Anantapur district. We will return to this stark difference in the analysis presented in Section 3.

All this evidence suggests that by having accurate predictions about the start of the monsoon, farmers could avoid costly mistakes. 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 at the 1 percent level. Thus, the later the onset, the lower will be accumulated rainfall during the cropping season.

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.[5] 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.

[5]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.

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 6 described in Section 2.4).

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 right plot on the top panel 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 (left graph), it only took 3 days for rainfall measured in Mahbubnagar to accumulate to its district average. The bottom panel of Figure 3 plots the number of days since June 1st that it takes each year for accumulated rainfall to reach the individual definition of the median respondent.

Two observations are in order. First, the start of the monsoon is far more volatile in Anantapur than in Mahbubnagar. Indeed, Figure 4 compares the cumulative density function of the onset of the monsoon arbitrarily defined as 20 mm (top panel) or 40 mm (bottom 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 context.

Second, at least visually, 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.[6]

When we regress the onset date (in days since June 1st) against a linear and quadratic 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).[7].

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 distribution of the onset of the monsoon from historical data. 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.

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.[8] This way we 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 constrain the respondents'

[6]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.

[7]When the individual definition is used, the number of observations increase dramatically and so estimated coefficients are more precise. But surprisingly, we find that while in Mahbubnagar the linear coefficient is positive and the quadratic is negative (both significant), the opposite is true in Anantapur (results not reported)

[8]Notice that the question did not specify the following year's monsoon, but rather the monsoon in general.

answers. One could argue 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. This suggests that predictions in Anantapur would be constrained and therefore less accurate. However, as we show below, respondents in Anantapur predict more accurately than in Mahbubnagar.[9]

Table 5 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 used between 3 and 4 bins. The historical distributions in Mahbubnagar seem less dispersed than the subjective ones, while the reverse is true in Anantapur. In "Moments and Properties of Distributions", we find that in Mahbubnagar respondents under-predict the mean and over-predict the variance, while the reverse is true in Anantapur. In sum, the first moment is better predicted in Anantapur while the second moment is better predicted in Mahbubnagar.

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,

[9]This method of elicitation has also been used successfully in Barrett et al. (2006) and Delavande and Kohler (2007), among others.

and S_k the number of stones from the available 10, 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]

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.[11]

Under "Accuracy", Table 6 reports the percentage (appropriately weighted) of individuals in each district whose subjective distribution is statistically different from the historical ones at different significance levels.

When the self-reported amount of rainfall needed to start sowing is used as definition, accuracy seems to be worse than when the average across respondents in the district is

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

[11]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 3.

used. It thus seems like bad forecasters not only miss the timing but also the quantum of rainfall needed to plant. The district average, however, compensates mistakes across respondents in such a way that it yields a historical distribution that is closer to the respondents' subjective 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. Thus, despite the evidence on the top and middle of Table 6 suggesting the contrary, respondents in Anantapur forecast better on average than those in Mahbubnagar.

There are several potential explanations that can account for this fact. First, if the correlation between rainfall in the gauge and rainfall in the village decreased with distance, rainfall gauges in Mahbubnagar could on average be further away from surveyed villages. While there is a mild positive correlation between the chi-square statistic and the distance from the rainfall gauge to the village, the relationship is not statistically significant. But more importantly, as reported in Table 1, the average distance from the village to the closest gauge is 4.14 miles in Mahbubnagar and 4.51 miles in Anantapur, so rainfall gauges are on average closer, not further away in Mahbubnagar. Another explanation could be that Mahbubnagar has worse quality of historical data. While it is true that some gauges have only 14 years of data while others have 35 years, on average there are 25.8 years of available data in Mahbubnagar but only 20.2 years in Anantapur. Again, there are no significant differences between both districts but if anything, Anantapur has fewer data points. Finally, there could be systematic differences in agricultural practices between districs that would make farmers in Anantapur more dependent on the monsoon than in Mahbubnagar. If we look at the area under irrigation during the monsoon season in both Anantapur and Mahabubngar, we find that indeed, while 27.61 percent of the land is irrigated in Mahbubnagar, this percentage drops to only 19.36 in Anantapur. [12]

[12]If we look instead at the number of farmers that use irrigation, the percentages are 17.13 in Mah-

Because irrigation can partially remedy mistakes made in planting decisions, farmers in Mahbubnagar have less to lose from a bad prediction than farmers in Anantapur. We will return to this issue in more detail in Section 3 when we assess differences in forecasting ability across respondents.

2.4 Summary statistics

Table 6 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 the table 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 (see Rosenzweig XX).

bubnagar and 15.75 in Anantapur.

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 and the value of the households' primary dwelling averaged Rs 69,000.

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, 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 6 also includes the percentage of land with different textures, the prevalence of intercropping and the slope and depth of the plots.

The bottom of Table 6 reports basic cropping patterns followed by farmers in our sample. Farmers in Anantapur tend to cultivate more plots (or divide plots into more subplots) and as a result they tend to grow more crops. As noted earlier, this may reflect the fact that weather in Anantapur is more erratic and thus farmers try to diversify more. Interestingly, most farmers tend 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.[13] The following section provides a rational explanation for this apparent behavior in Anantapur.

3 Empirical Analysis and Results

In the previous section we showed that the onset of the monsoon is more erratic in Anantapur than in Mahabubnagar. We now explore how farmers respond to a more volatile environment by focusing on the quantum of rainfall that farmers require to start planting and the percentage of total agricultural expenditures that farmers made before the start of the monsoon in 2006. 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 crop or variety that requires another type of land preparation.

The theory on irreversible alive investments predicts that the greater the uncertainty, the greater the value to wait so as to keep the option to invest. Thus, flexibility is more important the more uncertain the environment is. We would therefore expect that farmers in Anantapur require a *higher* quantum of accumulated rainfall to start planting and also that they purchase a *lower* fraction of agricultural inputs before the monsoon.

To test this, we run the following regression:

$$Y_{id} = \alpha D_d + X_{id}'\beta + \epsilon_{id},$$

[13]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.

where Y_{id} is either the minimum amount of rainfall that farmer i 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 and X_{id} are household level characteristics, including characteristics of the land the household cultivates. The coefficient of interest is α 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.

Table 8 reports the results of the above specification. Columns 1 and 4 regress the dependent variable against a district dummy, risk aversion and the discount, columns 2 and 5 control in addition for average land characteristics and columns 3 and 6 further control for other household level characteristics.

As the theory on irreversible investments would predict, the district dummy is positive and significant in columns (1)-(3) and negative and significant in columns (4)-(5). On average, farmers in Anantapur require an additional 5.2 millimeters of rain to start planting. According to the historical data, this amount of rainfall corresponds to waiting roughly 3 additional days to plant.

Likewise, farmers in Anantapur seem to not spend at all before the monsoon, indicating that they value being able to adapt the cropping patterns to the observed rainfall.

3.1 Who are the good predictors?

We now wish to compare the subjective distribution of the onset of the monsoon with the historical one, computed using both definitions for the onset of the monsoon.

The specification we run is the following:

$$\log(\chi^2_{iv}) = D_v + X'_{iv}\beta + \epsilon_{iv}$$

where χ^2_{iv} is the chi-square statistic computed according to the formula in (1) and that takes higher values the more the subjective and historical distribution of individual i in village (or district) v differ from each other. The variable D_v is a village (or a district) dummy and X_{iv} denotes a of household level characteristics that may affect accuracy. These variables are described in the Appendix, 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_{iv} would indicate that X_{iv} reduces accuracy, since higher χ^2_{iv} values are associated with larger differences between the 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. In columns 3 and 4, the historical distribution is computed using the individual definition. Columns 1 and 3 include a district dummy and the distance from the village to the rainfall gauge, while columns 2 and 4 include village dummies. While the district dummy is insignificant, the distance variable is significant and positive in column (3), indicating that basis risk is partly responsible for the observed lack of accuracy.

Variables that proxy for parameters in the utility function, as well as other demographic variables such as literacy, age and caste, do not seem to influence accuracy. Indeed the signs of the estimated coefficients are not robust across specifications. One notable exception is the age of the eldest household member, which is significant in columns (1) and (2) and of the expected sign: elder people improve accuracy, perhaps because their knowledge of the monsoon, accumulated over the years is transferred to the household head.

In addition, the wealthier the household as measured by the value of the primary dwelling, the worse the household is in forecasting the monsoon. 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. 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.

In sum, it appears as if better forecasters are precisely households whose incomes depend more heavily on the monsoon and households that lack proper risk-coping mechanisms to smooth weather shocks.

These results are compatible with a simple model where processing information involves a cost. Households acquire information until the benefit from better accuracy equal the costs of acquiring and processing the information. Thus, households for which the monsoon is important are willing to acquire more information and will on average be more accurate at predicting it.

3.2 Do farmers behave according to their expectations?

Finally, we wish to study whether respondent's behavior is consistent with their elicited subjective expectation. We compute two measures of when the household expects the monsoon to arrive. The first is simply the mean of the subjective distribution (Panel A of Table 9). The second is the area under the subjective distribution to the right of the historical mean in kartis (Panel B of Table 9). Both measures are correlated, but the second, which we call "pessimism", takes into account the shape of the subjective distribution by computing how much weight the household puts into a late monsoon arrival.

To test the correlation between expectations and decisions, we run the following set of regressions:

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

where Y_i are one-time decisions taken by individual i, μ_{Si} is the mean in kartis of the subjective distribution (the expected onset of the monsoon in kartis) and k_i is the actual onset of the monsoon computed using the "individual" or "district" definition. In other words, this variable in the odd-numbered columns takes on the kartis, for each farmer, in which accumulated rainfall for the relevant year recorded in the assigned rainfall gauge reached the minimum self-reported quantum of rainfall required to plant.[14] In even-numbered columns, the district average of individual definitions is used, rather than each individual definition, to compute the actual onset of monsoon.

In addition, the vector X_i includes household characteristics that may influence decision-making, in particular risk aversion and the discount rate, as well as the stratification variables. The coefficient of interest is again α. We expect it to be significant if expectations influence decision-making.

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).[15] The rainfall insurance paid off if accumulated rainfall until a predetermined date was below a certain trigger. Because the dates were determined in advance, households that expected the monsoon to start later would also find that the insurance policy more attractive, because the amount of time from the expected monsoon arrival until the fixed end-of-coverage date would be lower, thereby increasing the chances of a payout. In short, 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 9, 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. The coefficient on Pessimism in Panel B, columns (3) and (4) are also positive and significant.[16]

Second, we consider in columns (5)-(8) of Table 9 the average kartis in which the

[14]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.

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

[16]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 (2007).

household planted in 2004 and 2006, averaging across crops and plots cultivated. In this case, household that believe that the monsoon will arrive later should also plant later. Both the coefficient on the subjective mean (Panel A) and on "pessimism" (Panel B) are positive and significant. This results are significant because they provide direct evidence that *both* the minimum self-reported quantum of rainfall and its timing influence planting decisions. If farmers just waited for the minimum rainfall required to plant, irrespective of when the rains arrived, then expectations about the timing of the monsoon would be irrelevant. The coefficients of Panel A suggests that an increase of one kartis in the expectation would delay planting by roughly 0.4 kartis.

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 (from 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, expectations are a very powerful predictor of behavior, even more so than proxies of risk aversion or discount rates affecting the utility function. In addition, it seems that the simple mean of the subjective distribution is a better predictor than "pessimism". All in all these results are remarkable because they not only validates the elicitation method, but also indicate that heterogeneity in expectations more so than heterogeneity in preferences explain actual behavior.

4 Conclusions

In this paper we use unique survey data to assess how well farmers predict the start of the monsoon. We show that our elicitation method delivers consistent and informative subjective expectations that and that individuals behave according to them. 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. The results, therefore, underscore that fact that agents use a simple cost-benefit analysis model to acquire and process information.

Finally, we present some evidence that heterogeneity of expectations is critical to explain choice diversity, even more so that proxies for the parameters of the utility function. In sum, research should be directed towards incorporating expectations into decision-making processes.

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] 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.

[6] S. Gadgil and S. Gadgil. The Indian Monsoon, GDP and Agriculture. *Economic and Political Weekly*.

[7] 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.

[8] X. Giné, R. Townsend, and J. Vickery. Patterns of Rainfall Insurance Participation in Rural India. Working Paper, World Bank, 2007.

[9] 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.

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

[11] S. Klonner. Private Information and Altruism in Bidding Roscas. *The Economic Journal*, 2007.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] 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.

[19] P. Norris and R. Kramer. The Elicitation of Subjective Probabilities with Applications in Agricultural Economics. *Review of Marketing and Agricultural Economics*, 58(2).

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

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

[22] 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.

[23] 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.

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

[25] V. Zarnowitz and L.A. Lambros. Consensus and Uncertainty in Economic Prediction. *Journal of Political Economy*, 95(3):591–621, 1987.

Appendix

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
Literacy (1=Yes)	2004	Dummy variable equal to 1 if the household head is literate
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)
Slope greater than 1% (1=Yes)	2004	Pct of total acres cultivated by the household that have slope greater than 1%
Soil deeper than 80cm. (1=Yes)	2004	Pct of total acres cultivated by the household that are deeper than 80cm.

Bought insurance in 2004	2004	Dummy variable equal to 1 if the household bought rainfall insurance for 2004
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
Pct. of expenditures before onset of the monsoon	2006	Percentage of total inputs' expenditure invested before the onset of the monsoon in 2006
Subjective mean	2004	Mean kartis of the subjective distribution
Average planting kartis	2004 and 2006	Average kartis when planting took place across crops in relevant year
Actual onset of monsoon	2004 and 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.
Pessimism	2004	Area under the subjective distribution above the historical mean
Amount of rain (mms) required to start sowing	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.
HH replanted in last 10 years (1=Yes)	2006	Dummy variable equal to 1 if the household has ever replanted at least one crop

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 its 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 that were interviewed in the villages assigned to the rainfall gauge. "Distance in miles" is the average distance from the rainfall gauge among villages assigned to the rainfall gauge. "N. Years" is the average number of years with available data across rainfall gauges weighted by number of households in villages assigned to each rainfall gauge. The row "Total" reports the district averages of "Distance" and "N. Years" and district totals of "N. Villages" and "N. Households".

Table 2. Yield per acre of main subsistance crops

	Sorghum		Red Gram	
	(1)	(2)	(3)	(4)
Kartis	125.783	130.049	30.557	27.388
	[31.161]***	[33.314]***	[16.988]*	[18.788]
Kartis squared	-3.679	-3.834	-1.575	-1.421
	[0.871]***	[0.946]***	[0.615]**	[0.677]**
Controls	No	Yes	No	Yes
Mean dependent variable	154.265	154.081	26.231	26.260
Observations	249	248	892	891
R-squared	0.05	0.11	0.09	0.43

Notes: All regressions control for the variables used in the stratification.Other controls included in the regressions are plot level characteristics such as Soil is loamy sand (1=Yes), Soil is loam (1=Yes), Soil is clay loam (1=Yes), Slope > 1% (1=Yes), Soil is deeper than 80cm. (1=Yes) and Use of Improved or Hybrid Seeds (1=Yes). Each observation is a subplot. All regressions are estimated using OLS with clustering at the household level. Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Table 3. Kartis Codes

Code	Name	Dates in 2004
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. Expenditure before onset of monsoon in 2006

	Full Sample			Mahbubnagar			Anantapur		
	Total	Before Bharani	Before onset of Monsoon	Total	Before Bharani	Before onset of Monsoon	Total	Before Bharani	Before onset of Monsoon
Investments	Rs.	Pct.	Pct.	Rs.	Pct.	Pct.	Rs.	Pct.	Pct.
Bullock labour	1,745.5	15.9	25.9	2,242.3	18.6	30.4	768.41	0.00	0.00
Manual labour	3,768.5	12.1	20.4	4,090.5	16.9	28.4	3,135.1	0.00	0.00
Hiring tractor	2,875.8	14.1	33.7	3,486.6	17.3	40.6	1,674.5	0.01	0.05
Manure	1,702.3	51.7	64.7	1,674.2	77.5	90.1	1,757.4	0.03	0.17
Irrigation	876.40	35.5	46.8	1,008.6	46.5	59.4	616.38	0.00	0.06
Seeds (improved & hybrid)	2,560.7	4.90	25.6	2,287.0	8.01	42.1	3,099.0	0.00	0.02
Fertilizer	3,412.1	3.52	13.4	3,934.3	4.60	17.3	2,385.0	0.00	0.01
Total	16,941.3	15.2	28.4	18,723.5	20.5	37.4	13,435.8	0.01	0.04

Notes: Averages are weighted by the appropiate stratification weights.

Table 5: 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 ... bins 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	3.35	2.02	0.00	0.15	9.34	0.00	3.09	0.00	0.00
3 bins	43.94	30.25	26.10	33.75	63.02	40.00	46.19	0.37	0.00
4 bins	46.42	37.91	31.26	57.6	25.49	37.68	45.64	23.43	19.22
5 bins	5.80	13.94	27.06	8.18	1.35	22.33	5.08	30.55	35.94
6 bins	0.49	8.37	15.58	0.32	0.8	0.00	0.00	24.06	44.84
7 bins	0.00	6.28	0.00	0.00	0.00	0.00	0.00	18.04	0.00
8 bins	0.00	1.2	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	13.78	13.85	13.45	13.47	13.68	13.30	14.36	14.17	13.72
Standard Deviation	0.87	0.97	1.02	0.91	0.76	0.86	0.79	1.37	1.33
Unimodal (1=Yes)	93.13	70.35	77.83	92.10	96.22	93.13	95.06	21.89	49.11
Accuracy									
Subjective dist is different from historical at…									
1 percent		0.13	0.05	-	0.19	0.03	-	0.03	0.06
5 percent		0.26	0.12	-	0.32	0.12	-	0.13	0.12
10 percent		0.33	0.17	-	0.40	0.23	-	0.20	0.14
Subjective dist has same mode as historical (1=Yes)									
		0.43	0.43	-	0.51	0.22	-	0.30	0.55

Notes: Observations are weighted by the appropiate stratification weights.

Table 6. 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***
Other Household Characteristics							
Literacy (1=Yes)	0.37	0.48	0.00	1.00	0.33	0.46	0.133**
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
Use of intercropping system (1=Yes)	0.47	0.50	0.00	1.00	0.25	0.88	0.634***
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 [1]							
Number of Plots used	1.71	0.81	1.00	6.00	1.69	1.73	0.381***
Number of Crops cultivated	2.01	0.74	1.00	5.00	1.87	2.25	0.040
Farmer plants all crops in one kartis (1=Yes)	0.84	0.37	0.00	1.00	0.77	0.92	0.148***
Planting Kartis across crops	14.04	1.01	12.00	18.00	13.52	14.91	1.382***
Number of Observations	976				623	353	
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 teh most popular rainfed crops: sorghum, groundnut, castor, redgram and maize. 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 7. District differences

	Start millimeters			Pct. expenses before monsoon		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy Anantapur	4.243	5.076	5.153	-0.284	-0.284	-0.284
	[1.279]***	[0.810]***	[0.770]***	[0.012]***	[0.013]***	[0.013]***
Risk aversion	0.511	0.421	0.441	0.015	0.016	0.013
	[0.689]	[0.605]	[0.588]	[0.013]	[0.013]	[0.013]
Discount rate (*100)	0.374	0.198	0.495	-0.011	-0.011	-0.022
	[1.533]	[1.119]	[1.103]	[0.014]	[0.014]	[0.015]
Pct. of land with loamy sand soil		-2.928	-2.936		-0.081	-0.079
		[3.099]	[3.107]		[0.044]*	[0.044]*
Pct. of land with loam soil		8.343	8.556		-0.067	-0.070
		[3.410]**	[3.421]**		[0.041]	[0.042]
Pct. of land with clay loam		16.527	16.559		-0.065	-0.065
		[2.564]***	[2.595]***		[0.041]	[0.041]
Slope greater than 1% (1=Yes)		-1.286	-1.214		-0.005	-0.008
		[0.622]**	[0.621]*		[0.012]	[0.012]
Pct. land with expenditure in irrigation		1.234	1.201		0.019	0.022
		[1.464]	[1.475]		[0.028]	[0.032]
Soil deeper than 80cm. (1=Yes)		10.029	9.895		-0.052	-0.048
		[2.329]***	[2.400]***		[0.036]	[0.037]
Pct. of land with paddy		-3.944	-3.539		0.001	-0.009
		[2.383]	[2.386]		[0.040]	[0.042]
Forward caste (1=Yes)			1.052			-0.016
			[0.723]			[0.015]
Education			-0.203			-0.008
			[0.236]			[0.006]
Age of the household's head			0.033			-0.001
			[0.020]*			[0.000]***
Cultivated land in acres			0.053			-0.001
			[0.041]			[0.001]
Mean dependent variable	30.350	30.350	30.351	0.234	0.234	0.235
Observations	1,048	1,048	1,045	1,048	1,048	1,045
R-squared	0.04	0.47	0.48	0.45	0.45	0.46

Notes: In columns (1)-(3) the dependent variable is the amount of rain in milliliters. that the farmers require to start planting. In colums (4)-(6) the dependent variable is the percentage of agricultural expenditure made before the onset of the monsoon. All regressions control for the variables used in the stratification. All regressions are estimated using OLS with clustering at the village level. 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

	Average district		Individual	
	(1)	(2)	(3)	(4)
Dummy Anantapur	-0.125		-0.044	
	[0.104]		[0.100]	
Distance from village to rainfall gauge	-0.005		0.051	
	[0.023]		[0.023]**	
Risk aversion	-0.046	-0.059	0.029	0.015
	[0.067]	[0.067]	[0.042]	[0.039]
Discount rate	0.000	0.001	0.001	0.001
	[0.001]	[0.001]	[0.001]	[0.001]
Membership in BUA (1=Yes)	-0.444	0.16	-0.412	-0.011
	[0.076]***	[0.122]	[0.100]***	[0.107]
Weather info from informal sources (1=Yes)	-0.180	-0.094	-0.161	-0.101
	[0.085]**	[0.083]	[0.080]*	[0.077]
Market value of the house (Rs.100,000)	0.166	0.139	0.146	0.121
	[0.061]***	[0.054]**	[0.043]***	[0.031]***
Value of Owned land (Rs. 1.000,000)	0.013	-0.05	0.031	-0.025
	[0.045]	[0.029]*	[0.039]	[0.023]
Participation in chit fund (1=Yes)	0.257	0.126	0.244	0.159
	[0.103]**	[0.083]	[0.081]***	[0.072]**
Household is credit constrained (1=Yes)	-0.101	-0.067	-0.155	-0.137
	[0.064]	[0.065]	[0.048]***	[0.042]***
Per capita total income (Rs. 10,000)	0.014	0.012	0.006	0.004
	[0.009]	[0.005]**	[0.005]	[0.007]
Cultivated land in acres	-0.011	-0.001	-0.014	-0.004
	[0.005]*	[0.005]	[0.005]***	[0.004]
Pct. of land with paddy	0.300	0.362	0.222	0.144
	[0.208]	[0.188]*	[0.237]	[0.145]
Forward caste (1=Yes)	0.056	-0.018	0.09	0.077
	[0.091]	[0.096]	[0.078]	[0.059]
Literacy (1=Yes)	-0.058	-0.011	0.026	0.051
	[0.067]	[0.068]	[0.051]	[0.048]
Age of the household's head	0.001	0.001	0.031	0.015
	[0.003]	[0.003]	[0.022]	[0.023]
Age of the eldest household member	-0.007	-0.006	-0.002	-0.002
	[0.003]**	[0.003]**	[0.002]	[0.002]
Village Fixed Effects	No	Yes	No	Yes
Mean dependent variable	1.426	1.426	1.993	1.993
Observations	1,034	1,034	1,034	1,034
R-squared	0.080	0.18	0.12	0.27

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. All regressions are estimated using OLS with clustering at the village level and include village fixed effects in columns (2) and (4). 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. Do people behave according to their expectations?

	Bought insurance in 2004		Attended meeting and bought insurance in 2004		Average planting kartis in 2004		Average planting kartis in 2006		Pct. expenses before onset of the monsoon in 2006		Household replanted in last 10 years	
	District	Individual	District	Individual	District	Individual	District	Individual	District	Individual	District	Individual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Mean of subjective distribution												
Subjective mean of onset	-0.002	-0.003	0.080	0.077	0.396	0.391	0.392	0.400	-0.061	-0.062	-0.080	-0.082
	[0.003]	[0.003]	[0.034]**	[0.034]**	[0.082]***	[0.080]***	[0.105]***	[0.103]***	[0.013]***	[0.013]***	[0.021]***	[0.020]***
Actual onset of monsoon	0.009	0.010	0.057	0.065	0.136	0.160	0.837	0.773	-0.072	-0.062	-0.159	-0.163
	[0.006]	[0.006]	[0.042]	[0.040]	[0.043]***	[0.035]***	[0.138]***	[0.111]***	[0.028]**	[0.023]**	[0.075]**	[0.066]**
Risk aversion	-0.023	-0.024	-0.104	-0.111	-0.098	-0.105	-0.082	-0.104	0.019	0.021	0.001	0.004
	[0.008]***	[0.009]***	[0.055]*	[0.058]*	[0.063]	[0.063]	[0.098]	[0.093]	[0.017]	[0.017]	[0.049]	[0.046]
Discount Rate (*100)	-0.018	-0.019	-0.118	-0.123	-0.007	-0.013	-0.022	-0.104	-0.061	-0.053	0.056	0.075
	[0.014]	[0.014]	[0.079]	[0.078]	[0.124]	[0.123]	[0.213]	[0.216]	[0.026]**	[0.026]**	[0.050]	[0.046]
Pct. of land with paddy	0.073	0.073	-0.291	-0.291	-0.495	-0.486	-0.864	-0.929	0.142	0.150	0.075	0.085
	[0.043]	[0.042]*	[0.190]	[0.179]	[0.210]**	[0.202]**	[0.337]**	[0.338]***	[0.048]***	[0.049]***	[0.121]	[0.124]
Household is credit constrained (1=Yes)	-0.009	-0.009	0.004	-0.004	-0.018	-0.025	-0.032	-0.031	0.003	0.001	0.100	0.094
	[0.013]	[0.013]	[0.084]	[0.080]	[0.083]	[0.082]	[0.114]	[0.113]	[0.016]	[0.016]	[0.042]**	[0.043]**
Observations	57333	57333	5043	5043	1033	1033	917	917	739	739	666	666
R-squared	0.02	0.02	0.12	0.12	0.16	0.17	0.17	0.18	0.28	0.28	0.13	0.13
Panel B: Pessimism												
	Bought insurance in 2004		Attended meeting and bought insurance in 2004		Average planting kartis in 2004		Average planting kartis in 2006		Pct. expenses before onset of the monsoon in 2006		Household replanted in last 10 years	
	District	Individual	District	Individual	District	Individual	District	Individual	District	Individual	District	Individual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Pessimism	-0.005	-0.003	0.175	0.228	0.753	0.845	0.687	0.947	-0.138	-0.164	-0.132	-0.197
	[0.019]	[0.019]	[0.094]*	[0.093]**	[0.200]***	[0.205]***	[0.259]**	[0.297]***	[0.025]***	[0.026]***	[0.057]**	[0.055]***
Actual onset of monsoon	0.008	0.009	0.072	0.090	0.207	0.249	1.035	1.052	-0.110	-0.109	-0.199	-0.222
	[0.007]	[0.006]	[0.042]	[0.041]**	[0.053]***	[0.046]***	[0.173]***	[0.163]***	[0.028]***	[0.024]***	[0.076]**	[0.069]***
Risk aversion	-0.023	-0.024	-0.106	-0.111	-0.100	-0.109	-0.082	-0.106	0.021	0.023	0.003	0.006
	[0.008]**	[0.009]**	[0.049]**	[0.053]**	[0.064]	[0.064]*	[0.104]	[0.098]	[0.017]	[0.016]	[0.050]	[0.047]
Discount Rate (*100)	-0.020	-0.021	-0.104	-0.106	0.040	0.026	0.027	-0.089	-0.071	-0.058	0.036	0.064
	[0.014]	[0.015]	[0.078]	[0.077]	[0.130]	[0.129]	[0.213]	[0.216]	[0.024]***	[0.024]**	[0.052]	[0.047]
Pct. of land with paddy	0.074	0.074	-0.299	-0.289	-0.648	-0.622	-0.980	-1.011	0.154	0.159	0.100	0.102
	[0.044]	[0.042]*	[0.189]	[0.176]	[0.247]**	[0.230]**	[0.379]**	[0.365]***	[0.052]***	[0.051]***	[0.122]	[0.126]
Household is credit constrained (1=Yes)	-0.009	-0.009	-0.010	-0.017	-0.044	-0.053	-0.050	-0.043	0.010	0.007	0.110	0.102
	[0.013]	[0.013]	[0.080]	[0.076]	[0.083]	[0.082]	[0.120]	[0.117]	[0.016]	[0.015]	[0.041]**	[0.043]**
Mean dependent variable	0.290		0.454		14.179		14.211		0.249		0.300	
Observations	57333	57333	5043	5043	1033	1033	917	917	739	739	666	666
R-squared	0.02	0.02	0.12	0.11	0.13	0.12	0.16	0.15	0.27	0.26	0.13	0.12

Notes: In Panel A, "Subjective mean of onset" is the mean in kartis of the subjective distribution elicited from each respondent. In Panel B, "Pessimism" is the area under the subjective distribution to the right of the mean historical distribution. "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). All regressions are estimated using OLS with clustering at the village level. Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Figure 1. District Location in Andhra Pradesh, India

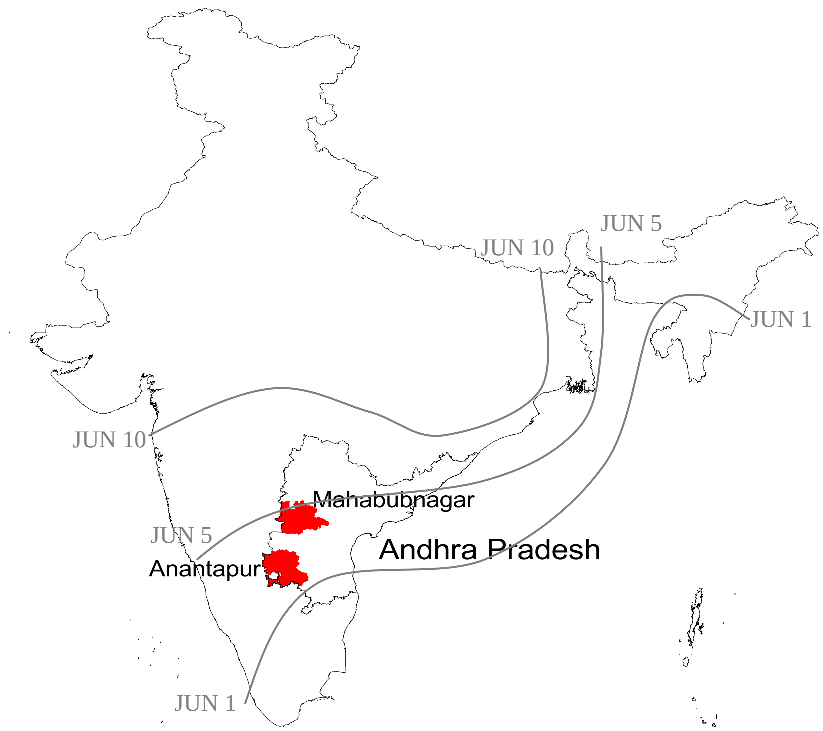

Figure 2. Village Location in Mahbubnagar

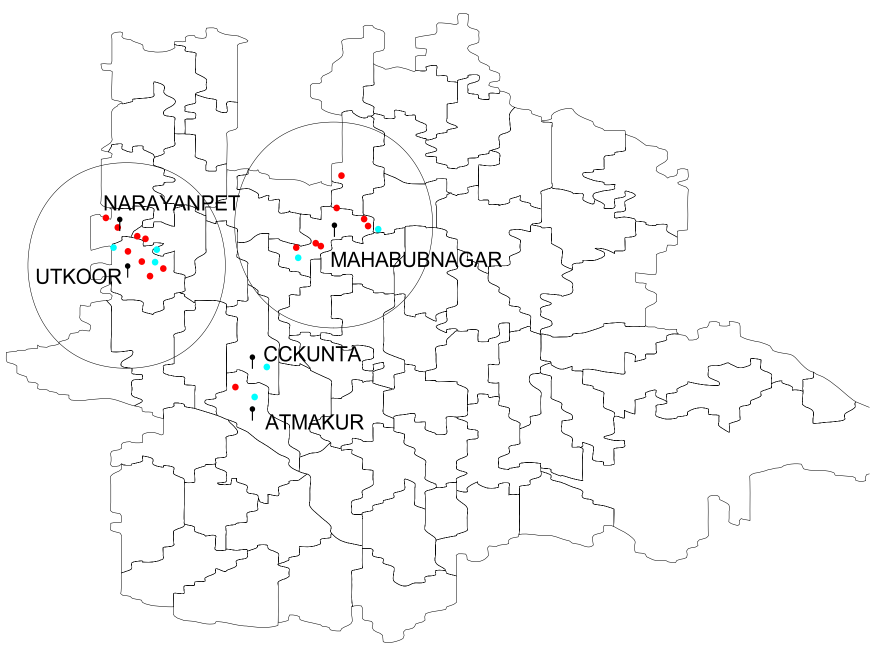

Figure 3. Onset of the Monsoon over Time

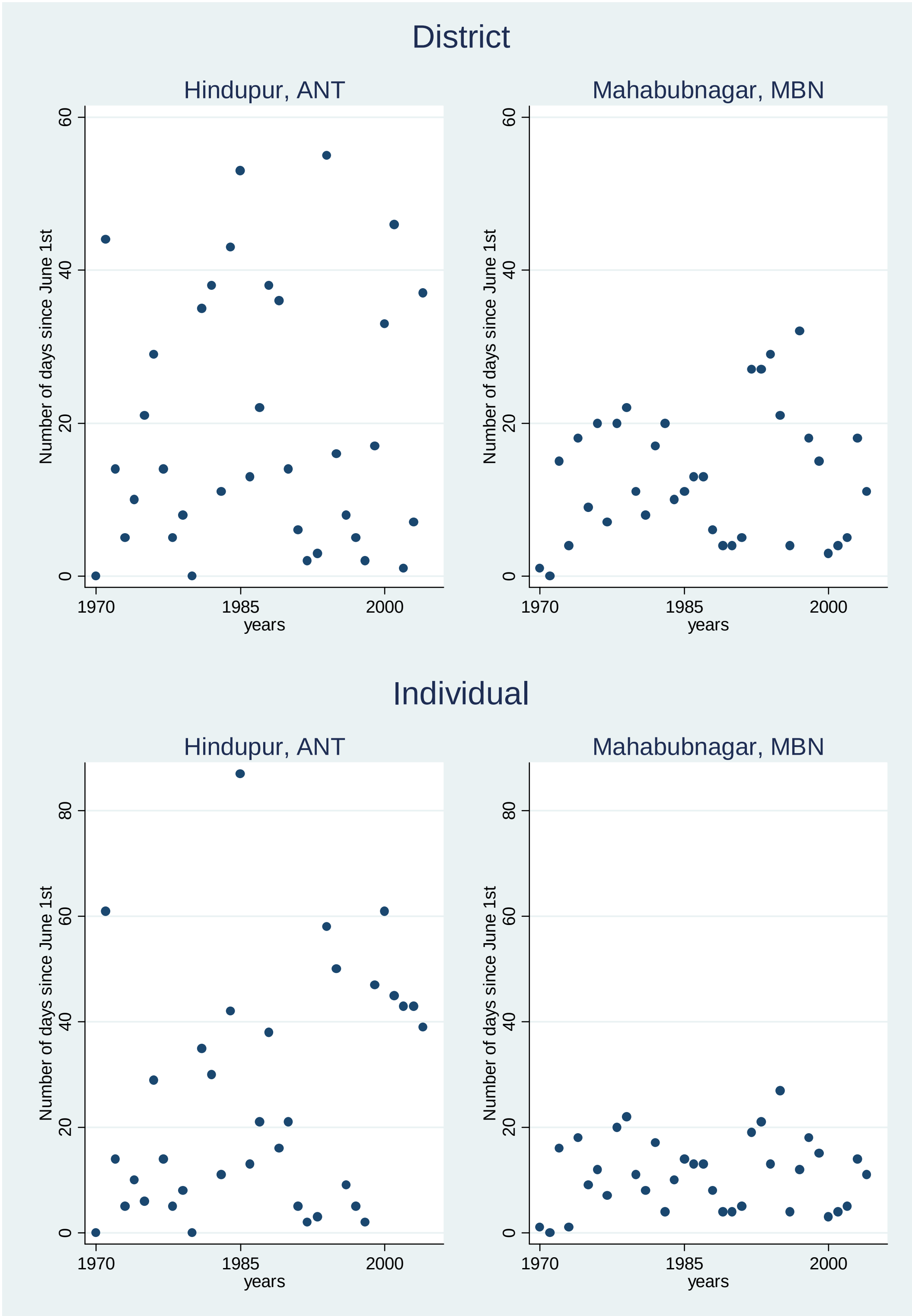

Figure 4. Cumulative Distribution Functions