Forecasting When it Matters: Evidence from Semi-Arid India

Xavier Giné* | Robert M. Townsend | James Vickery
World Bank | MIT | FRB of NY

This draft: October 2015

Abstract

Decision-making under uncertainty depends not only on preferences, endowments and constraints but also on subjective probabilities. We elicit from a sample of farmers their beliefs about the timing of the onset of the monsoon, which signals the optimal time to plant. We assess the accuracy of beliefs by comparing them to historical data and find substantial heterogeneity in beliefs and accuracy. Elicited beliefs explain differences in observed behavior consistent with the predictions of a simple model in which information is costly to acquire: poorer farmers whose income is more dependent on the monsoon are more accurate.

Keywords: Information Acquisition, Expectations, Forecasting, Climate.
JEL Codes: D81, D84, O13, Q54.

*Email addresses: xgine@worldbank.org, rtownsen@mit.edu, and james.vickery@ny.frb.org.
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 computing rainfall absorption of different soil types. We also thank Jean Marie Baland, Abhijit Banerjee, Shawn Cole, Adelaine Delavande, Quy-Toan Do, Pascaline Dupas, Michael Kremer, Lance Lochner, Hanan Jacoby, David McKenzie, Mark Rosenzweig, Chris Udry and participants in several conferences and seminars for valuable comments. Paola de Baldomero Zazo, Francesca de Nicola, Joan Pina Martí and Laxman Timilsina 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.

Uttara chusi, yattara gampa.

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

Telugu Proverb

1 Introduction

Decision-making under uncertainty depends on preferences and the subjective probabilities that agents attach to the different states. The literature has typically assumed that these beliefs coincide with the true probabilities and so differences in behavior across individuals facing the same constraints and endowments must be driven by differences in preferences. If one allows for differences in both probability distributions and preferences, however, then a fundamental identification problem arises, since behavior can be explained by differences in preferences, beliefs or both (Manski, 2004).

In this paper we elicit the subjective distribution of the monsoon onset, an event of critical importance to lives of individuals engaged in rainfed agriculture in semi-arid India. In these areas, weather risk is a major source of income fluctuations. Rosenzweig and Binswanger (1993), for example, find that the delay of the monsoon can have considerable negative effects on agricultural yield and profits.

To cope with risk, households rely on ex-ante and ex-post strategies that typically trade expected profits for lower risk (Walker and Ryan, 1990; Morduch, 1995). One such strategy, particularly relevant to our purposes, is the choice of an optimal time to plant that coincides with the onset of the monsoon (Fein and Stephens, 1987; Rao et al., 2000 and Gadgil et al., 2002).

Because the onset varies from year to year, deciding when to plant involves significant judgment. If farmers plant after the first rains, but subsequent rains are scattered and fall several days apart, the seeds may not germinate. Farmers are then forced to either replant or to abandon the crop altogether, both resulting in significant losses. Alternatively, being too conservative by postponing planting until one is certain that the monsoon has arrived is also costly, because yield will typically be lower (Bliss and Stern, 1982; 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 these 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.

Our data provide direct evidence that misjudging the actual arrival of the monsoon is costly. In 2006, about a quarter of our sample had replanted in the past and a full 73 percent had abandoned the crop at least once in the prior 10 years due to poor rains. In addition, the extra expenses required to replant are large, equivalent to 20 percent of total average production expenses. Thus, it seems that farmers would benefit greatly from having accurate beliefs about the onset of the monsoon.

Our survey covers 1,050 farming households living in 37 villages in southern India. Experimental methods are 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 reports the minimum depth of soil moisture that he would require to start planting. This measure is then converted into a quantum of rainfall using the absorption capacity of the respondent's specific soil type. The subjective calendar distribution is obtained by giving each respondent 10 stones and a sheet of paper with boxes corresponding to 13-14 day calendar periods. The respondent is then instructed to distribute the 10 stones across the different boxes according to the likelihood that the monsoon starts in the calendar period indicated by each box (see Delavande et al. 2011a for a review of these methods).

This elicitation method delivers subjective expectations that are quite heterogeneous, even within villages. To understand the implications of this heterogeneity, we develop a simple model that combines heterogeneity in beliefs, irreversible investment and costly information acquisition (as in, for example, Reis, 2006).

Consistent with the data, the model predicts that in villages where the monsoon onset is more volatile, farmers plant later. Intuitively, if initial rains are not very informative about the onset of the monsoon, the option value of waiting for more information before planting is higher.

Second, farmers who believe the monsoon will start later are also more likely to plant later and to be less likely to replant. All of these findings, consistent with the model, provide strong evidence that individuals use their subjective expectations to make decisions, even after controlling for self-reported proxies of risk aversion, discount rates and the actual start of the monsoon.[1]

The model also predicts that the accuracy of farmers' beliefs are dictated by a simple cost-benefit analysis of information acquisition. We find that accurate farmers are poorer, with more rainfall dependent income and more likely to be credit constrained. Put differently, accurate farmers are less able to cope with weather risk and have livelihoods that are more dependent on the monsoon. Thus, accuracy is rationally explained by how relevant the event is to the forecaster (Brunnermeier and Parker, 2005), rather than by heuristics or "rules of thumb" (Kahneman, Slovic and Tversky, 1982; Rabin, 1998 or DellaVigna, 2009 for a review).

Finally, we show that inaccuracy about the onset of the monsoon is costly. In particular, an improvement in our measure of accuracy from the 25th to the 75th percentile of the distribution would increase gross agricultural income by 8 to 9 percent. Similarly, accurate farmers are less likely to replant, thus suggesting fewer planting mistakes. Of course, farmers that do make mistakes tend to have less at stake because their incomes are less depend on the monsoon.

More generally, it is important to distinguish behavior driven by preferences from behavior driven by beliefs because they may result in different policy prescriptions. If farmers have accurate beliefs, and planting choices are driven by differences in preferences,

[1] We are not the first to show the predictive power of elicited subjective expectations. In other contexts, Bernheim and Levin (1989), Lochner (2007), van der Klaauw and Wolpin (2008), Delavande (2008), Mahajan et al. (2011) and Giné and Jacoby (2015) all show that elicited beliefs predict observed behavior.

then policies that distort those choices could be welfare reducing. In contrast, if differences in planting choices are the result of inaccurate beliefs, then policies that lower the cost of acquiring information could enhance welfare. In addition, because differences in behavior can be explained by differences in expectations independently of differences in preferences, these results advocate for the collection and use of expectations data to predict behavior.

Our results suggests that in a context where information is costly, interventions that only disclose information about an event may be ineffective because individuals that care about the event will have already gathered the relevant information while individuals that do not care about it will not pay any attention to it. This may provide a rational for the mixed evidence found in the literature about the impacts of information campaigns.[2]

More broadly, this paper contributes to a growing literature that measures expectations and its impacts on various outcomes (see Manski, 2004 for an excellent review; Norris and Kramer, 1990 for an early review of elicitation methods applied to agricultural economics or more recently, Attanasio, 2009; Delavande et al. 2011a and Delavande, 2013 that also review the literature in developing countries and advocate the collection of expectations data).[3]

While this literature shows that beliefs are sometimes systematically biased (for example, DellaVigna and Malmendier 2006 show that people signing up for gym memberships are overconfident about future attendance; Allcott 2013 shows that consumers under-

[2]See for example Allcott (2011) and Ferraro et al. (2011) on energy and water conservation and Dupas (2009) and Ashraf et al. (2013) on health-related products.

[3]Bernheim and Levin (1989) study how beliefs of social security benefits affects consumption, savings, and retirement decisions. Similarly, Dominitz and Manski (2007) uses beliefs on portfolio returns, van der Klaauw and Wolpin (2008) uses beliefs in life expectancy and Dominitz (1998) uses beliefs in future income to study similar decisions. Delavande (2008) studies how takeup of preventive birth control measures depend on expectations of their efficacy and Mahajan et al. (2011) examine the adoption of antimalarial insecticide-treated bednets in India. Lochner (2007) shows how perceived arrest probabilities correlate with individual characteristics and with future criminal behavior. Dominitz and Manski (1996), Jensen (2010), and Delavande and Kohler (2012) study perceived returns to schooling and educational choices, while McKenzie et al. (2013) study perceived returns to migration. Luseno et al. (2003) and Lybbert et al. (2007), study the extent to which cattle herders in Kenya update their priors on rainfall expectations in response to new information. Carroll (2003) and Mankiw, Reis, and Wolfers (2003) study inflation expectations, showing that consumers disagree and that individuals beliefs tend to lag those of professional forecasters.

estimate vehicle gas mileage and Jensen (2010) and McKenzie et al. (2013) show that individuals underestimate the returns to schooling and migration, respectively), we are the first ones to combine historical and survey data to assess individual *heterogeneity* in accuracy and its impacts on welfare.

The rest of the paper is organized as follows. The next section presents a simple model and a set of testable predictions, Section 3 describes the context and the data collected, Section 4 takes the model predictions to the data and finally Section 5 concludes.

2 Theory

Consider a risk-neutral farmer that after receiving some rain at the beginning of the monsoon season must decide whether to immediately plant a rainfed crop or to wait. The benefit of waiting is that the farmer can make a more informed decision later, which is valuable because if the rains do not signal the onset of the monsoon and are thus not followed by more rains, the seeds planted may not germinate. The cost of waiting, however, is that planting immediately yields higher agricultural output if the monsoon has indeed arrived.

The farmer has a prior about when the monsoon will arrive which may differ from the objective probability, but may invest cognitive resources in either updating the prior or another activity that does not depend on the monsoon.

In what follows, we first lay out the assumptions and we then characterize the optimal planting strategy, the expected profits from growing the rainfed crop, and the decision to update the prior. The Online Appendix contains more details about the model and the proofs of the propositions.

Timing of events. The timing of events is shown in Figure 1. There are three periods: 0, 1 and 2. The monsoon can arrive at the end of period 1 or in period 2. If the monsoon arrives at the end of period 1, the state s is "Monsoon" ($s = M$), while if it arrives in period 2, the state s is "Delay" ($s = D$).

At the beginning of period 1 the farmer observes an initial amount of rain $r \in [0, R]$ that is informative of the state s in that state is correlated with the rain (see below). In period 0 the farmer decides whether to forgo additional income and update the prior that the state is $s = M$ using the signal r received. The farmer then chooses to either plant the rainfed crop or to wait. We assume that planting in period 2 is also profitable, and so farmers that did not plant in period 1, will do so in period 2. At the end of period 2, the farmer receives harvest income which depends on the state and the timing of planting.

Updating of Priors. Farmers have priors $\widehat{\mu}$ of the state $s = M$ that may differ from the objective probability μ. Let $\nu = |\widehat{\mu} - \mu|$ be a proxy for accuracy such that farmers for whom ν is larger will have more inaccurate priors.[4]

Income. If the farmer decides to plant in period 1, he obtains net income $\overline{y}$ if the monsoon arrives at the end of period 1 ($s = M$). If the farmer waits until period 2 to plant, he obtains net income $\underline{y}$, irrespective of the state. We assume that $\overline{y} > \underline{y} > 0$, reflecting the cost of waiting to plant in period 2 when the state is $s = M$ ($\overline{y} > \underline{y}$) and that planting in period 2 is still profitable ($\underline{y} > 0$). If the farmer planted in period 1 but the monsoon was delayed ($s = D$), the farmer can still plant in period 2. We normalize the combined net income from planting and failing in period 1 and replanting in period 2 to 0. Finally, if attention is not paid in updating the prior in period 0, the farmer obtains additional income y_0 from an alternative activity.

Correlation between signal r and state s. The probability density function of observing signal r, conditional on state s, is $f(r, \delta|s)$ which is continuous in both arguments and satisfies $f(r, 0|D) = f(r, 0|M)$, for all $r \in [0, R]$, $f(R, 1|D) = 0$ and $f(0, 1|M) = 0$.

[4]The model considers only one period and focuses on the decision to update the prior or not given that the farmer has prior $\widehat{\mu}$. In a multi-period version of the model where each period is subdivided into the three sub-periods corresponding to the periods 0, 1 and 2, farmers that in some period t decided not to update the prior, would reach the same decision in all subsequent periods. More formally, if priors were not updated in period t, then $\widehat{\mu}_t = \widehat{\mu}_{t-1}$. Alternatively, if the prior were updated in some period t, then it would always be updated and would converge to the true probability μ.

In addition,

$$\frac{\partial f(r,\delta|M)}{\partial r} \geq 0 \quad \text{and} \quad \frac{\partial f(r,\delta|D)}{\partial r} \leq 0 \quad (=0 \quad \text{if} \quad \delta = 0),$$

because $f(r, 0|s)$ is a constant. The parameter $\delta \in [0, 1]$ measures the degree to which signal r is informative, with $\delta = 0$ indicating that the signal is completely uninformative.

2.1 Optimal Planting Decision

Consider first a farmer that decides not to update the prior in period 0. The farmer will plant if the expected gains from planting are positive. The gains from planting when priors are not updated Ω is the difference in expected utility between planting and waiting in period 1:

$$\Omega = Ey(\text{Plant}) - Ey(\text{Wait}) = \Pr(s = M)\overline{y} + \Pr(s = D)0 - \underline{y} = \widehat{\mu}\overline{y} - \underline{y}. \tag{1}$$

Next, consider a farmer that updates her prior of the onset in period 0. The difference in expected utility between planting and waiting in period 1 now depends on the signal r received and informativeness δ. We can write $\Omega_r(\delta)$ as

$$\begin{aligned} \Omega_r(\delta) &= Ey(\text{Plant}|r) - Ey(\text{Wait}|r) \\ &= \Pr(s = O|r)\overline{y} - \underline{y} \\ &= \frac{\overline{y}}{1 + \phi(r,\delta)\frac{1-\widehat{\mu}}{\widehat{\mu}}} - \underline{y} \end{aligned} \tag{2}$$

where $\phi(r, \delta) = \frac{f(r,\delta|D)}{f(r,\delta|M)}$.

Given the assumptions made, $\phi(r, 0) = 1$ and $\Omega_r(0) = \Omega$, for all signals $r \in [0, R]$. Intuitively, this suggests that when the signal is uninformative ($\delta = 0$), there are no benefits to updating the prior.

Let μ^* be the probability of state $s = M$ such that $\mu^*\overline{y} = \underline{y}$. We assume that the objective probability μ satisfies $\mu < \mu^*$.[5]

Here we consider the case where the prior satisfies $\widehat{\mu} < \mu^*$ while the Online Appendix considers the case where the prior satisfies $\widehat{\mu} \geq \mu^*$. When the prior satisfies $\widehat{\mu} < \mu^*$, a farmer that does not update in period 0 will always wait in period 1 because $\Omega < 0$. In this case it is easy to show that $\Omega_R(\delta)$ ($\Omega_0(\delta)$) is increasing (decreasing) in δ.[6]

In addition, there exists a level of informativeness $\delta_R^* < 1$ such that $\Omega_R(\delta_R^*) = 0$.[7] This cutoff level of informativeness δ_R^* is important because if informativeness satisfies $\delta < \delta_R^*$, then it is always optimal to wait in period 1, regardless of the signal r received. If however $\delta > \delta_R^*$, then there exists a cutoff signal $r_\delta < R$ such that the farmer will wait if $r < r_\delta$, but will plant otherwise. This cutoff r_δ satisfies $\Omega_{r_\delta}(\delta) = 0$. Thus, signal informativeness plays an important role in the optimal planting strategy, as stated in our first proposition.

Proposition 1 *If $\widehat{\mu} < \mu^*$, the amount of rain required to plant in period 1 is (weakly) decreasing in signal informativeness δ. More formally, if $\delta > \delta_R^*$, then $\frac{\partial r_\delta}{\partial \delta} < 0$.*

Figure 2 plots the function $\Omega_r(\delta)$ for $r = R, r = 0$ and $r = r_{\delta'}$ when $\widehat{\mu} < \mu^*$ (and thus $\Omega = \Omega_r(0) < 0$). The optimal planting strategy for a farmer with $\delta = \delta' > \delta_R^*$ is to plant if signal $r \in [r_{\delta'}, R]$ and to wait if signal $r \in [0, r_{\delta'})$. If in contrast $\delta < \delta_R^*$, then as mentioned before, the optimal planting strategy is to wait regardless of signal r.

Since the expected gains from planting in period 1 depend positively on beliefs $\widehat{\mu}$, it also follows that farmers with higher posterior beliefs $\widehat{\mu}$ are more likely to plant in period 1 as stated in the next proposition.

[5] From the definition of Ω in expression (1) we have that if $\widehat{\mu} = \mu^*$, then $\Omega = 0$.

[6] In equation (2), informativeness δ only enters the function $\Omega_r(\delta)$ through its dependency on $\phi(r, \delta)$. In addition, $\Omega_r(\delta)$ depends negatively on $\phi(r, \delta)$. The assumptions about $f(r, \delta|s)$ ensure that $\phi(R, \delta)$ is decreasing and $\phi(0, \delta)$ is increasing. Thus $\Omega_R(\delta)$ ($\Omega_0(\delta)$) is increasing (decreasing) in δ as was to be shown.

[7] By assumption $f(R, 1|D) = 0$, therefore $\phi(R, 1) = 0$ and $\Omega_R(1) = \overline{y} - \underline{y} > 0$. Since the prior satisfies $\widehat{\mu} < \mu^*$, the difference in expected utility when $r = R$ and $\delta = 0$ is negative, $\Omega_R(0) < 0$. By Bolzano's theorem, the continuous function $\Omega_R(\delta)$ with values of opposite signs in the interval $\delta \in [0, 1]$ has a root, and therefore δ_R^* exists. By continuity, since $\Omega_R(1) > 0$ and $\Omega_R(\delta_R^*) = 0$, $\delta_R^* < 1$.

Proposition 2 *The probability of planting in period 1 is (weakly) increasing in beliefs* $\widehat{\mu}$ *of state* $s = M$*.*

2.2 Expected Profits

In this section we derive expressions for expected profits using both the priors $\widehat{\mu}$ and the actual probability μ. We then evaluate the loss in expected profits from having inaccurate priors.

Expected profits from growing the rainfed crop for a farmer with prior $\widehat{\mu}$ that decides not to update it are denoted $\widehat{E}Y_0$, where $\widehat{E}(\cdot)$ indicates that the expectation is taken over the prior $\widehat{\mu}$. Expected profits are thus

$$\widehat{E}Y_0(\widehat{\mu}) = \begin{cases} \underline{y} & \widehat{\mu} < \mu^* \\ \widehat{\mu}\overline{y} + (1-\widehat{\mu})0 = \widehat{\mu}\overline{y} & \widehat{\mu} \geq \mu^*. \end{cases} \tag{3}$$

Expected profits EY_0 for the same farmer with prior $\widehat{\mu}$ but taking expectations over the true probability μ of state $s = M$ (instead of the prior $\widehat{\mu}$) are

$$EY_0(\widehat{\mu}) = \begin{cases} \underline{y} & \widehat{\mu} < \mu^* \\ \mu\overline{y} & \widehat{\mu} \geq \mu^*. \end{cases} \tag{4}$$

We note that the farmer still decides to wait if $\widehat{\mu} < \mu^*$ or to plant if $\widehat{\mu} \geq \mu^*$ but in this case expected profits are $\widehat{\mu}\overline{y}$ in expression (3) and $\mu\overline{y}$ in expression (4).

Similarly, when $\widehat{\mu} < \mu^*$ expected profits $\widehat{E}Y_a$ when priors are updated as a function of beliefs $\widehat{\mu}$, can be written as

$$\widehat{E}Y_a(\widehat{\mu}) = \begin{cases} \underline{y} & \delta < \delta_R^*(\widehat{\mu}) \\ \underline{y} + \int_{\widehat{r}_\delta}^{R} \widehat{\pi}_r \Omega_r(\widehat{\mu})\, dr & \delta \geq \delta_R^*(\widehat{\mu}) \end{cases} \tag{5}$$

where the dependency of δ_R^* and Ω_r on beliefs $\widehat{\mu}$ is made explicit. As defined earlier, signal $\widehat{r}_\delta$ is such that $\Omega_{\widehat{r}_\delta}(\widehat{\mu}, \delta) = 0$, and $\widehat{\pi}_r = f(r, \delta|M)\widehat{\mu} + f(r, \delta|D)(1-\widehat{\mu})$.

Expected profits EY_a evaluated at the true probability μ of state $s = M$ can be written as

$$EY_a(\widehat{\mu}) = \begin{cases} \underline{y} & \delta < \delta_R^*(\widehat{\mu}) \\ \underline{y} + \int_{\widehat{r}_\delta}^{R} \pi_r \Omega_r(\mu)\, dr & \delta \geq \delta_R^*(\widehat{\mu}) \end{cases}$$

where in this case Ω_r and π_r are evaluated at the true probability μ.

As before, expected profits $EY_a(\widehat{\mu})$ evaluated at μ still depend on $\widehat{\mu}$ through the cutoff signal $\widehat{r}_\delta$ and $\delta_R^*(\widehat{\mu})$. For example, a farmer with prior $\widehat{\mu} > \mu^*$ will face critical values $\widehat{r}_\delta < r_\delta$ and $\delta_R^*(\widehat{\mu}) < \delta_R^*(\mu)$ and so will be more likely to plant when he should have waited.

We now assess whether individuals with less accurate priors tend to have lower expected profits. We compute the difference in expected profits $EY_a(\mu) - EY_a(\widehat{\mu}) \geq 0$ both evaluated at the true probability μ and show that it increases with accuracy. We state this result in the form of a proposition.

Proposition 3 *The more inaccurate the prior of state $s = M$ is, the lower are expected profits from rainfed crops. More formally,*

$$\frac{\partial \left[EY_a(\mu) - EY_a(\widehat{\mu}) \right]}{\partial \nu} \geq 0$$

where $\nu = |\widehat{\mu} - \mu|$ as defined before.

In words, individuals with more inaccurate priors will tend to have lower expected profits evaluated at the true probability μ because they are more likely to plant when they should have waited and are more likely to wait when they should have planted. To be clear, we consider these actions a "mistake" because the decision is made *before* the state is revealed and because they would be different if priors were accurate.[8]

[8] Ex-post errors made by farmers with accurate priors such as planting when in fact the monsoon was delayed ($s = D$), or waiting when the monsoon actually arrived in period 1 ($s = M$) are not considered "mistakes" because these actions were optimal given the (accurate) information available.

2.3 Choice of Attention

The farmer decides whether to invest cognitive resources in updating the prior by solving the following problem:

$$\max\left\{\widehat{E}Y_0(\widehat{\mu}) + y_0, \widehat{E}Y_a(\widehat{\mu})\right\}$$

The farmer thus updates the prior if $\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0$. If $\widehat{\mu} < \mu^*$, from expressions (3) and (5) we have

$$\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) = \begin{cases} 0 & \delta < \delta^*_R(\widehat{\mu}) \\ \int_{\widehat{r}_\delta}^{R} \widehat{\pi}_r \Omega_r(\widehat{\mu})\, dr & \delta \geq \delta^*_R(\widehat{\mu}) \end{cases} \tag{6}$$

It is clear from expression (6) (and expression (8) in the Online Appendix) that the decision to update the prior will depend on the returns to the alternative activity y_0 and informativeness δ as stated in the next proposition.

Proposition 4 *The less relevant the monsoon is to overall net income, the less accurate will posterior beliefs be. In addition, the more informative the signal is, the more accurate will posterior beliefs tend to be. More formally,*

$$\frac{\partial Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial y_0} < 0 \quad and \quad \frac{\partial Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial \delta} \geq 0.$$

Thus, farmers with access to irrigated crops whose net income do not depend on accurately predicting the timing of the monsoon will tend to have less accurate beliefs. Similarly, if the signals are not very informative it does not pay to update the prior and farmers will tend to have more inaccurate priors. There is therefore a form of complementarity between signal informativeness and the decision to update the priors.

3 Data and Context

We now describe the household survey data and historical rainfall data used to test the propositions of the model in the previous Section. Survey data were collected using a specialized survey that was also designed to evaluate a rainfall insurance pilot program described in Giné, Townsend and Vickery (2008) and Cole et al. (2013). The survey took place after the 2004 harvest, and covered 1,051 households in 37 villages from two drought prone and relatively poor districts in the states of Andhra Pradesh and Telangana, India. The same households were resurveyed in 2006, 2008, 2009 and 2010, so a few variables are constructed using these later rounds.[9]

Survey villages are assigned to the closest rainfall gauge, typically located less than 5 miles away. There are five rainfall gauges in Mahbubnagar district and five in Anantapur district. Table 1 presents the number of villages and households assigned to each rainfall gauge, the average distance from each gauge to the villages assigned to it and the number of years with available daily rainfall data.

Table 2 presents summary statistics of the respondents' household characteristics, their land characteristics and of other variables used in the analysis.[10] The average household head is 47 years old and has been farming for 25 years. Thirty percent of households have a member older than 60 whose experience in farming could be used by the head in updating his priors about the onset.[11] The head of household has an average of 3.3 years of schooling and a literacy rate (measured by the self-reported ability to read and write) of only 37 percent. Respondents cultivate on average 6.28 acres valued at Rs. 246,000 (USD 5,234) and the value of the households' primary dwelling averages Rs 69,000 (USD

[9]The sampling framework used a stratification based on whether the household attended a marketing meeting and whether the household purchased a deficit rainfall insurance policy in 2004. Within each strata, households were randomly selected.

[10]Appendix Table A1 contains a detailed description of how variables are constructed. All statistics in Table 2 are weighted by the sampling weights. See Giné, Townsend and Vickery, 2008 for further details.

[11]Rosenzweig and Wolpin (1985) develop a theory that explains the predominance of intergenerational family extension, kin labor and the scarcity of land sales from the fact that elders have experiential knowledge about farming practices specific to the plots of land owned by the household.

1,468), thus confirming that the sample consists mostly of smallholder farmers.

Table 2 also includes the percentage of land with different textures, the prevalence of intercropping and the slope and depth of the top soil of their plots.[12] The bottom of Table 2 reports basic cropping patterns followed by our sample of farmers. Interestingly, most farmers plant all of the rain sensitive crops at once, instead of planting different plots at different times as a way to diversify risk. They do so because of the relatively high fixed costs of land preparation and the prevalence of intercropping that requires that crops be sown at the same time.

The main cropping season runs from June to November coinciding with the monsoon. Farmers grow several of cash and subsistence crops that vary in their yield's sensitivity to drought. The main cash crops grown in the area are castor (mostly in Mahbubnagar, covering 34 percent of the cultivated area), groundnut (covering half of the cultivated area in Anantapur) and paddy (grown in both districts and covering 9 percent of cultivated area in the sample). The main subsistence crops are grams and sorghum (covering less than 10 percent each). All crops with the exception of paddy are rainfed (less than 5 percent of plots are irrigated). Paddy is almost exclusively irrigated (84 percent of all plots use groundwater irrigation only available to wealthier farmers).

Table 2 reports that farmers in Anantapur tend to plant later, around July 6 while farmers in Mahbubnagar plant around June 15th. Thus, farmers in Anantapur plant much later than the normal onset date of June 3rd according to the Indian Meteorological Department (IMD), drawn in gray in Figure 3.[13]

According to our survey data, between 1996 and 2006, twenty-two percent of house-

[12] Intercropping is frequently used in the semi-arid tropics and 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 soil depth.

[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.

holds replanted at least once, and almost three quarters had abandoned the crops 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) that accounted for 23 percent of total production costs.

All in all, this evidence suggests that farmers could avoid costly mistakes by having accurate beliefs about the start of the monsoon.[14]

3.1 Onset of Monsoon

We use two definitions of onset of the monsoon. The first definition is based on 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.[15] We label this quantum of rainfall from each respondent, the "individual definition" of the onset. The second, or "district definition" is simply the average of the individual definitions just described for each district. The average onset of the monsoon in millimeters is 28.95 mm in Mahbubnagar and 33.05 mm in Anantapur.[16]

[14]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 (see also Rosenzweig and Binswanger, 1993 and Rosenzweig and Udry, 2014).

[15]The calculations for rainfall penetration for various soil textures use 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.

[16]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

Figure 4 shows that there is volatility in the onset of the monsoon. It plots for a representative rainfall station in each district, the time that it took each year for daily rainfall to accumulate to the district definitions described above, 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 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 definition.

To be sure, by the time farmers decide when to plant, they may have received numerous signals about the likelihood that the monsoon will arrive in the coming days. If upon receiving such a signal, farmers knew with certainty the onset and therefore the optimal time to plant, then beliefs about the onset would be inconsequential to the planting decision, since they would simply wait for the conclusive signals. But a long and rich folk tradition of trying to predict the arrival of the monsoon rains indicate that these signals are noisy at best and that the *prior* distribution still matters as it affects the farmers' *conditional* expectation as assumed in the model of Section 2.[17,18]

Indeed, Figure 5 shows that the onset of the monsoon conditional on the IMD forecast still displays substantial volatility.[19] For a given year, it plots the difference in days between the actual onset of the monsoon plotted in Figure 3 and the forecast from the IMD.

the soil being red. See "Land Characteristics" in Table 2.

[17]This accumulation of indigenous knowledge over thousands of years is reflected 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 (Fein and Stephens, 1987).

[18]In terms of the model in Section 2, the prior distribution is $\widehat{\mu}$ and the conditional expectation is $\Pr(s = M|r_j)$.

[19]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 3 to the forecast over Kerala. Rosenzweig and Udry (2014) use the long-range rainfall forecast of the monsoon and show that in places where the forecast is more accurate agricultural investment decisions react more to the forecast.

Figures 4 and 5 show that there is no trend in the start of the monsoon over time, consistent with Goswami et al. (2006) that 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.[20,21]

The lack of a time trend in the onset of the monsoon suggests that one can weight each year equally when constructing the historical distribution.

In addition, both the onset of the monsoon in Figure 4 and the conditional onset of Figure 5 are more volatile in Anantapur than in Mahbubnagar (p-val of equal means is 0.000). In terms of the model of Section 2, it appears that signal informativeness in Anantapur district tends to be lower than in Mahbubnagar. Proposition 1 of the model suggests that farmers in this more erratic environment will respond by requiring more rain to start planting, a prediction we test in Section 4.

3.2 Consistency of Subjective Distributions

The subjective distribution of the monsoon arrival is obtained experimentally by giving the respondent 10 stones and a sheet of paper with boxes corresponding to the different 13-14 day periods called "kartis".[22] The respondent is instructed to place all the stones in the different boxes according to his estimate of the likelihood that the monsoon would start in each period. The question does not specify any year in particular and should be interpreted as the respondent's prior distribution of the onset of the monsoon.

[20] 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.

[21] 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).

[22] Kartis or naksatra in Sanskrit are based on the traditional solar calendar (Fein and Stephens, 1987; Gadgil et al., 2002) and are used by farmers to determine the timing of agricultural activities.

The simple model of Section 2 considers only 2 possible dates of monsoon arrival and thus the prior distribution is a single number $\widehat{\mu}$. In practice, the monsoon can arrive anytime from early June to mid July and thus a multi-date support is required. More importantly, with the full distribution over a multi-period support we can use moments higher than the mean to assess behavior.

This method of elicitation has been successfully used elsewhere (Giné and Jacoby, 2014 for a recent example and Delavande et al. 2011a for a review). The fact that every respondent provided a subjective distribution and no respondent allocated less than the total 10 stones given also gives us confidence that the elicitation was successful. In addition, no respondent allocated a stone in the first or last period of the predefined support, indicating that it did not constrain the respondents' answers.[23]

Table 3 provides further evidence of the consistency of subjective distributions as compared to the historical ones, computed both under the "individual" and "district" definitions. The highest and lowest bin in the support of both the subjective and historical distributions are remarkably similar, except perhaps in Anantapur, where respondents seem to underestimate the range. In addition, most respondents put stones in 3 or 4 different bins, and no respondent puts all 10 stones in a single bin, suggesting perceived uncertainty about the onset.

The model of Section 2 is silent as to whether $\widehat{\mu} < \mu$ or viceversa. In their theory of optimal expectations, Brunnermeier and Parker (2005) suggest that because individuals derive utility from an early monsoon, they would tend to predict that the onset would arrive earlier ($\widehat{\mu} > \mu$ according to the model). We do not find evidence that supports this prediction as the mean of the subjective distribution is lower than that of the historical distribution using either definition. If anything, the subjective mean is higher, indicating that on average individuals expect the monsoon to arrive later. There is however some

[23] One concern is that respondents are constrained by the limited number of stones given. Delavande et al. (2011b) compares the accuracy of eliciting distributions with 10 or 20 stones and concludes that the number of stones does improve accuracy but only when the number of support points with positive mass is large.

evidence of over-confidence since the standard deviation of the subjective distribution is lower than that of the historical distribution under either definition (p-value is 0.000 in both cases).

3.3 Accuracy

The proxy for accuracy in the model of Section 2 is $\nu = |\widehat{\mu} - \mu|$. In our multi-period support distribution, we measure accuracy of beliefs in the data by computing a simple chi-square statistic. 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

$$\chi^2 = \sum_k \frac{(\sqrt{S/H}H_k - \sqrt{H/S}S_k)^2}{H_k + S_k}, \tag{7}$$

where

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

and k indexes the kartis. 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 to the respondent) minus 1.[24]

We find that around 28 percent of households report a subjective distribution that is significantly different at the 10% level from the historical one computed using the "individual" definition.[25]

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

[25] The percentage is 26 when the historical distribution is computed with the "district" definition and there is no statistical difference between the two percentages.

4 Empirical Analysis and Results

In this section we formally test the propositions of the model in Section 2.

4.1 Uncertainty and irreversible investments

In Figure 5 described in Section 3, we showed that the conditional onset of the monsoon is more erratic in Anantapur district than in Mahabubnagar. Proposition 1 of the model states that in places with more uncertainty, or equivalently, with less informative signals, farmers wait longer before incurring irreversible investments. We consider two such investments: the amount of rainfall required to plant and the purchase of inputs before the monsoon arrives.

Both decisions have a certain 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 for lack of subsequent rains. 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 Proposition 1, we run the following regression:

$$Y_{iv} = \alpha V_{iv} + X'_{iv}\beta + \delta_v + \epsilon_{iv},$$

where Y_{iv} is the investment decision of farmer i in village v, V_{iv} is the standard deviation of the conditional onset of the monsoon plotted in Figure 5, that is, the difference in days from the IMD forecast until the onset of monsoon, X_{iv} are household level characteristics and δ_v are village dummies.

The key coefficient of interest is α. When the dependent variable is the minimum amount of rainfall required to plant we expect α to be positive as households living in more volatile areas will require more rain to start planting. Similarly, when the dependent variable is the percentage of expenditure before the monsoon we expect α to be negative as households facing more volatility will wait for the actual onset to purchase the agricultural inputs.

Table 4 reports the results. All regressions include land, household controls and the stratification variables. The dependent variable in columns (1)-(3) is the amount of rain in millimeters that respondents require to start planting and in columns (4)-(6) it is the percentage of all agricultural expenditures made before the onset of the monsoon. Columns (1)-(2) and (4)-(5) use the individual definition of onset while columns (3) and (6) use the district definition to compute the standard deviation of the conditional onset of the monsoon V_{iv}.

Consistent with Proposition 1, the proxy for uncertainty is positive and significant in columns (1)-(3) and negative in columns (4)-(6) (albeit only significant in column 4). The coefficient on risk aversion is positive, as expected, but not precisely estimated. According to column 2, an increase of one standard deviation in volatility V_{iv} is equivalent to waiting 3.6 additional days to plant and to reducing input expenditures before the monsoon by 1.7 percent.

Importantly, these results also confirm that the prior $\widehat{\mu}$ indeed satisfies $\widehat{\mu} < \mu^*$, at least on average, even though they can take on any value in the unit interval. Since one can show that the amount of rain required to plant would be *increasing* in signal informativeness when $\widehat{\mu} \geq \mu^*$, the fact that the rain required to plant decreases in signal informativeness suggests that $\widehat{\mu} < \mu^*$.

4.2 Do beliefs affect behavior?

Proposition 2 states that the probability of planting early is increasing in the belief of an early onset. We test this proposition more generally by assessing whether the behavior of respondents is influenced by their beliefs. We use the mean of the subjective distribution as a proxy for the household's beliefs of the timing of the monsoon onset. [26] To test the correlation between beliefs and decisions, we estimate the following set of regressions:

$$Y_{iv} = \alpha\mu_{Siv} + \beta r_{iv} + \delta k_{iv} + X'_{iv}\gamma + \delta_{iv} + \epsilon_{iv},$$

where Y_{iv} are production decisions taken by individual i in village v, μ_{Siv} is the mean in kartis of the subjective distribution (the expected onset of the monsoon in kartis) and α is the coefficient of interest. The variable r_{iv} is the individual definition of the monsoon in millimeters (minimum amount of rainfall required to start planting) and is included to isolate differences in the expected timing of the monsoon from differences in the individual definition. The variable k_{iv} is the actual onset of the monsoon computed using the individual definition.[27] We include it to isolate the influence of the subjective distribution on behavior, because it captures all the information available to the farmer when planting decisions are made. Put differently, we assess the relative importance of the unconditional vis à vis the conditional expectation. The vector X_{iv} includes household characteristics that may influence decision-making, in particular risk aversion and the discount rate. We would expect that risk averse individuals would be more inclined to purchase insurance, to plant later, to replant less and to make fewer purchases before the monsoon. Similarly, individuals that discount more heavily the future would also tend to plant later and make fewer purchases before the monsoon as these costly activities are pushed into the future. Finally, we include village dummies δ_{iv} to focus on within village variation in behavior.

We consider four production decisions Y_{iv}, some of which are measured in multiple rounds. First, whether the household bought rainfall insurance in 2004 (conditional on having attended a marketing meeting) in column (1).[28] The insurance policy paid out if

[26]Alternatively, we compute the probability that the monsoon will start before a given date using the subjective probability. The results are similar to those in Table 5 (not shown).

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

[28]Insurance policies were not advertised outside the marketing meeting, and therefore most farmers

accumulated rainfall during the coverage period was below a certain trigger. Because the end date of the policy coverage period was predetermined, households that believe that the monsoon would start late would also believe that the probability of a payout was higher as there would be fewer rainy days during the coverage period. The prediction therefore is that households that expect a later onset, would be more likely to purchase insurance. Second, we consider the average kartis in which the household planted in 2004, 2006, 2008 and 2009, averaging across crops and plots cultivated. Proposition 2 predicts that households that expect a late onset of the monsoon will also plant later. Third, we consider in column (3) whether the household has replanted in 2006 or in any of the previous 10 years. The prediction here is that households that believe that the monsoon will start later should have a lower probability of replanting, because by the time they plant the rains have likely arrived. Finally we look in column (4) at the percentage of expenditures before the actual onset of the monsoon in 2006 and 2009. Again, households that believe that the monsoon will start later will make fewer purchases before the monsoon actually sets in, although they would have liked to purchase inputs in advance to be prepared for the onset of the monsoon.

The results are reported in Table 5 and provide strong evidence that the prior distribution of the monsoon onset is a powerful predictor of behavior in the expected direction. According to column (1), individuals that believe that the monsoon will arrive late are more likely to purchase the insurance product.[29] The number of observations in column (1) is lower because it only includes individuals that attended the insurance marketing meeting conducted only in a subset of villages. In column (2) the coefficient implies that an increase of one kartis in the mean of the subjective distribution of the monsoon onset would delay planting by roughly 0.4 kartis or approximately one week. In column (3) it suggests that an increase of one kartis in the mean of the beliefs distribution reduces the

that did not attend the meeting would not have heard about the insurance product.

[29] Notice that risk aversion is significant but negative, implying that more risk-averse households are *less* likely to purchase insurance. This contradictory result is discussed in detail in Giné, Townsend and Vickery (2008).

probability that the household has replanted in 2006 or in any of the previous 10 years by 5.4 percent. In column (4) the coefficient has the expected sign but is not statistically significant. As reported in Table 4 columns (4) and (5) the variation in the percentage of expenditures variable comes from differences across villages rather than differences within villages and therefore the inclusion of village dummies wipes out most of the variation.[30]

It is worth noting that if insurance markets were perfect within a village and households collectively diversified risk by varying the timing of planting then accuracy would not be predictable and the inclusion of village fixed effects in Table 5 would wipe out the correlation between beliefs and behavior. However, since accuracy is predictable, even with village fixed effects we are able to reject the null of complete within-village insurance markets.

Finally, while the prior distribution has explanatory power, the proxies for risk aversion and the discount rate do not explain behavior except in column (1). All in all, these results are remarkable because they not only validate the elicitation method, but also indicate that heterogeneity in beliefs, more so than heterogeneity in preferences, explain actual behavior.

4.3 Who has accurate beliefs?

We now turn to test 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 depends on how relevant the production of rain-sensitive crops is to overall income and on how informative the signals are. The specification we run is the following:

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

[30]Similarly to Table 4 column (4) the coefficient of interest is negative and statistically significant when village dummies are not included.

where χ^2_{iv} is the chi-square statistic computed according to the formula in (7) for individual i in village v. A higher value of χ^2_{iv} indicates less accurate beliefs relative to the historical distribution. The vector Z_v contains village-level variables such as the proxy for uncertainty described in the previous subsection and the distance from village v to the rainfall gauge. The vector X_{iv} includes a set of household level characteristics that may affect accuracy. These variables are described in Section 3, and include basic demographics, such as caste, literacy and education, wealth, the 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 6 reports the results. In columns (1)-(2), the historical distribution is computed using the district average of the individual definitions of the onset of monsoon while in columns (4)-(5), the individual definition is used. Columns (1) and (4) include village covariates, while columns (2) and (5), our preferred specification, include village dummies. Columns (3) and (6) report the p-value of an F-test that variables are jointly insignificant for the specification that includes village dummies.

The standard deviation of the conditional historical distribution is positive and significant in column (1) confirming Proposition 3 in the model, but not in column (4). Distance to the rainfall gauge is significant and positive in columns (1) and (5), 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 inaccuracy of beliefs.

Variables that proxy for parameters in the utility function do not seem to influence accuracy. The lowest p-value of an F-test that variables 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 related to cognition per se because both literacy and 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 0.19).[31]

More interestingly, 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 recalling the monsoon arrival. An F-test that both wealth variables are jointly insignificant is always rejected at the 10 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 column (7) but not in column (3). 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 rainfed crops are also of the right sign and mostly significant. The F-test is always rejected at the 10 percent level.

In sum, confirming Proposition 3 of the model, households whose incomes depend more heavily on the monsoon and households that lack proper risk-coping mechanisms to smooth weather shocks predict better the onset of the monsoon.

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

4.4 Does Accuracy Matter?

According to Proposition 4 of the model, inaccuracy increases the scope for mistakes in the timing of planting and so farmers with less accurate beliefs experience lower yields of rainfed crops and should be more likely to replant.

To test Proposition 4 we first estimate the following yield and profit regression

$$\begin{aligned} y_{cit} &= \alpha_0 \log(\chi_i^2) + \alpha_{NP} \log(\chi_i^2) \times D_{NP} + D_c + D_t + D_i + \\ &\quad D_{NP} \times D_t + \log(\chi_i^2) \times D_t + X_{ic}'\beta + \epsilon_{cit}, \end{aligned} \tag{8}$$

[31] The only household characteristic that appears significant is the Herfindahl index of soil type. Because the minimum moisture required to plant (ie the individual definition of the onset) depends on the soil type, farmers with different soil types will have a harder time reporting the correct definition and so will tend to be inaccurate.

where y_{cidt} is the yield per acre of crop c grown by farmer i in year t or total profits of crop c grown by farmer i in year t, where profits are computed by subtracting labor costs from the value of production. The variables D_c is a crop c dummy, D_{NP} is a dummy that takes value 1 if the crop is rainfed, that is, if the crop is not paddy (No Paddy 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 (7), and X_c is a set of plot characteristics aggregated by crop and individual, that includes soil characteristics and the use of intercropping, among others. We use crop yields for 2004 and 2006, the years for which we have data.[32] Since we have multiple crop observations per households we include household fixed effects and we cluster standard errors at that level.

Second, we estimate the same specification of Table 5 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 whether the household has replanted in 2006 or any of the prior 10 years and consumption in 2004, 2006 and 2007. The variable of interest is α_χ.

Panel A of Table 7 reports the results from regression (8). In columns (1) and (2) we find that lower accuracy results in lower yields for rainfed crops relative to paddy yields. As discussed, paddy is an irrigated crop and thus its cultivation is less subject to the vagaries of the monsoon. Because the coefficient in the variable $\log(\chi_i^2)$ is hard to interpret, we compute the increase in agricultural production for an individual with average landholdings growing rainfed crops, if accuracy were increased from the 25th percentile of the distribution to the 75th percentile.[33] We find an average gain between Rs 1,500 and Rs 1,800 depending on the 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. It is perhaps remarkable that we find income gains

[32] In later data collection rounds yields at the plot level are not available.

[33] 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.

with only two years of data. In the regression however we exploit geographical variation in the timing of the monsoon onset, accuracy and timing of planting from multiple plots per households. Columns (3) and (4) use profits as the dependent variable. We find the coefficient of interest to be less precisely estimated.

Columns (1)-(4) of Panel B presents strong evidence that more accurate beliefs are correlated with a lower probability of having replanted in the past. If accuracy increases from the 25th percentile of the distribution to the 75th percentile, the probability of replanting falls by 5.3 percent or 5.8 percent depending on the definition of onset used. In columns (5)-(8) of Panel B, we find that more accurate beliefs are also correlated with higher consumption levels but the coefficients are not statistically significant.

5 Discussion and Conclusions

The results indicate that accuracy of beliefs vary across respondents, although the costs of acquiring information appear low given that the onset of the monsoon is publicly observable. Since accuracy is negatively correlated with wealth, perhaps farmers that appear less accurate simply had higher opportunity cost of their time and did not put as much effort in answering the questionnaire, including the questions that elicited the subjective distribution. We test this hypothesis by correlating our proxy of inaccuracy $\log(\chi_i^2)$ with an index of correct answers to abstract questions about rainfall insurance. Because the correct answer to these questions also required the attention of the respondent, one would observe a negative correlation between inaccuracy and knowledge of insurance if inaccuracy were indeed related to their opportunity cost of time. As it turns out, columns (1) and (2) of Table 8 report a weak negative correlation (as expected) but one that is far from statistical significance. We thus conclude that our proxy for accuracy is not driven by the opportunity cost of time of the respondent or his lack of effort answering the survey.

Since inaccurate farmers are more likely to make planting mistakes as evidenced by their higher probability of replanting and their lower yields, an obvious question that immediately arises is why they do not imitate more successful neighbors in their planting and investment decisions. This strategy however presupposes that inaccurate farmers *know* who the better forecasters are. Columns (3) to (6) of Table 8 show that inaccurate farmers are typically considered worse farmers by their peers and more importantly, in columns (7) and (8) that they are less likely to predict who the good rainfed crop producers are.[34] The variable of interest is again $\log(\chi_i^2)$ and the coefficient tends to be negative and significant, suggesting that inaccurate farmers (with high χ^2) are less likely to be considered either progressive farmers or good rainfed farmers and are also less likely to identify farmers with rainfed yields in the top 25th percentile.

Finally, while the data support the assumption that $\widehat{\mu} < \mu^*$, and $\mu < \mu^*$, the average prior $\widehat{\mu}$ could be higher or lower than the true probability μ. If $\widehat{\mu} < \mu$ then farmers that do not update and have thus inaccurate priors would tend to delay planting relative to those that are accurate as their expectations of $s = M$ are low. Columns (9)-(10) of Table 8 regress the average planting time against our measure of inaccuracy or χ^2 and find that indeed inaccurate farmers (high χ^2) tend to plant later, suggesting that $\widehat{\mu} < \mu$.

In sum, we find evidence that individuals with stronger incentives to gather information about the monsoon appear to have more accurate beliefs, consistent with a simple model of costly information acquisition. Thus 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 individual.

Our elicitation method delivers informative prior subjective expectations that are a powerful predictor of behavior, even more so than proxies for parameters in the utility function and thus, we conclude that eliciting expectations data is important and should be incorporated in the analysis of decision-making problems.

[34] In columns (3) and (4) the dependent variable is a dummy for whether the respondent considers himself a progressive farmer. Columns (5) and (6) the dependent variable is the number of times a respondent is cited as being a good farmer of rainfed crops. In columns (7) and (8) the dependent variable is the number of times a respondent is cited as being a good farmer, so long as average yields of rainfed crops in 2004 and 2006 were in the top 25th percentile.

References

[1] H. Allcott. Social norms and energy conservation. *Journal of Public Economics*, 95(9):1082–1095, 2011.

[2] H. Allcott. The Welfare Effects of Misperceived Product: Data and Calibrations from the Automobile Market. *American Economic Journal: Economic Policy*, 5(3):30–36, 2013.

[3] N. Ashraf, B. K. Jack, and E. Kamenika. Information and subsidies: complements or substitutes? *Journal of Economic Behavior and Organization*, 88:133–139, 2013.

[4] O. Attanasio. Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review*, 99(2):87–92, 2009.

[5] B. D. Bernheim and L. Levin. Soical Security and Personal Savings: An Analysis of Expectations. *American Economic Review*, 79(2):97–102, 1989.

[6] C. J. Bliss and N. H. Stern. *Palanpur, the Economy of an Indian Village*. Oxford University, Oxford, England, 1982.

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

[8] C. D. Carroll. Macroeconomic Expectations of Households and Professional Forecasters. *Quarterly Journal of Economics*, 118(1):269–98, 2003.

[9] S. Cole, X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. Barriers to Household Risk Management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1):104–135, 2013.

[10] A. Delavande. Pill, Patch, or Shot? Subjective Expectations and Birth Control Choice. *International Economic Review*, 49(3):999–1042, 2008.

[11] A. Delavande. Probabilistic Expectations in Developing Countries. *National Bureau of Economic Research*, 6:1–20, Forthcoming. First published online as a Review in Advance on December 20, 2013.

[12] A. Delavande, X. Giné, and D. McKenzie. Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence. *Journal of Development Economics*, 94:151–163, 2011a.

[13] A. Delavande, X. Giné, and D. McKenzie. Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, 26(3):479–497, 2011b.

[14] A. Delavande and H. P. Kohler. The Impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi. *Demography*, 49(3):1011–36, 2012.

[15] S. DellaVigna. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2):315–372, 2009.

[16] J. Dominitz. Earnings Expectations, Revisions, and Realizations. *The Review of Economics and Statistics*, 80(3):374–88, 1998.

[17] J. Dominitz and C. F. Manski. Eliciting Student Expectations of the Returns to Schooling. *Journal of Human Resources*, 31(1):1–26, 1996.

[18] P Dupas. What matters (and what does not) in households' decision to invest in malaria prevention? . *American Economic Review*, 99(2):224–230, 2009.

[19] M. Fafchamps. Sequential Labor Decisions Under Uncertainty: an Estimable Household Model of West African Farmers. *Econometrica*, 61(5):1173–1197, 1993.

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

[21] P. J. Ferraro, J. J. Miranda, and M. K. Price. The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–322, 2011.

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

[23] 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.

[24] X. Giné and H. Jacoby. Markets, Contracts, and Uncertainty: A Structural Model of a Groundwater Economy. World Bank, 2015.

[25] 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.

[26] 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, 2006.

[27] R. Jensen. The (Perceived) Returns to Education and the Demand for Schooling. *Quarterly Journal of Economics*, 125(2):515–48, 2010.

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

[29] L. Lochner. Individual Perceptions of the Criminal Justice System. *American Economic Review*, 97(1):444–60, 2007.

[30] 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.

[31] T. Lybbert, C. Barrett, J. McPeak, and W. Luseno. Bayesian herders: asymmetric updating of rainfall beliefs in response to external forecasts. *World Development*, 35(3):480–497, 2007.

[32] A. Mahajan and A. Tarozzi. Time Inconsistency, Expectations, and Technology Adoption: The Case of Insecticide Treated Nets. U.C. Berkeley, 2011.

[33] 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.

[34] N. G. Mankiw, R. Reis, and J. Wolfers. Disagreement about Inflation Expectations. *National Bureau of Economic Research*, 18:209–70, Macroeconomics Annual 2003. edited by Mark Gertler and Kenneth Rogoff.

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

[36] 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? *Journal of Development Economics*, 102:116–127, 2013.

[37] J. Morduch. Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives*, 9(3):103–114, Summer 1995.

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

[39] 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.

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

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

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

[43] R. Reis. Inattentive Producers. *Review of Economic Studies*, 73:793–821, 2006.

[44] 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.

[45] M. Rosenzweig and C. Udry. Forecasting Profitability. Yale U., 2014.

[46] M. Rosenzweig and K. Wolpin. Specific Experience, Household Structure, and Intergenerational Transfers: Farm Family Land and Labor Arrangements in Developing Countries. *Quarterly Journal of Economics*, 100(Supplement):961–987, 1985.

[47] 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.

[48] W. van der Klaauw and K. Wolpin. Social Security and the Retirement and Savings Behavior of Low-Income Households. *Journal of Econometrics*, 145(1–2):21–42, 2008.

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

Figure 1. Timeline of Model

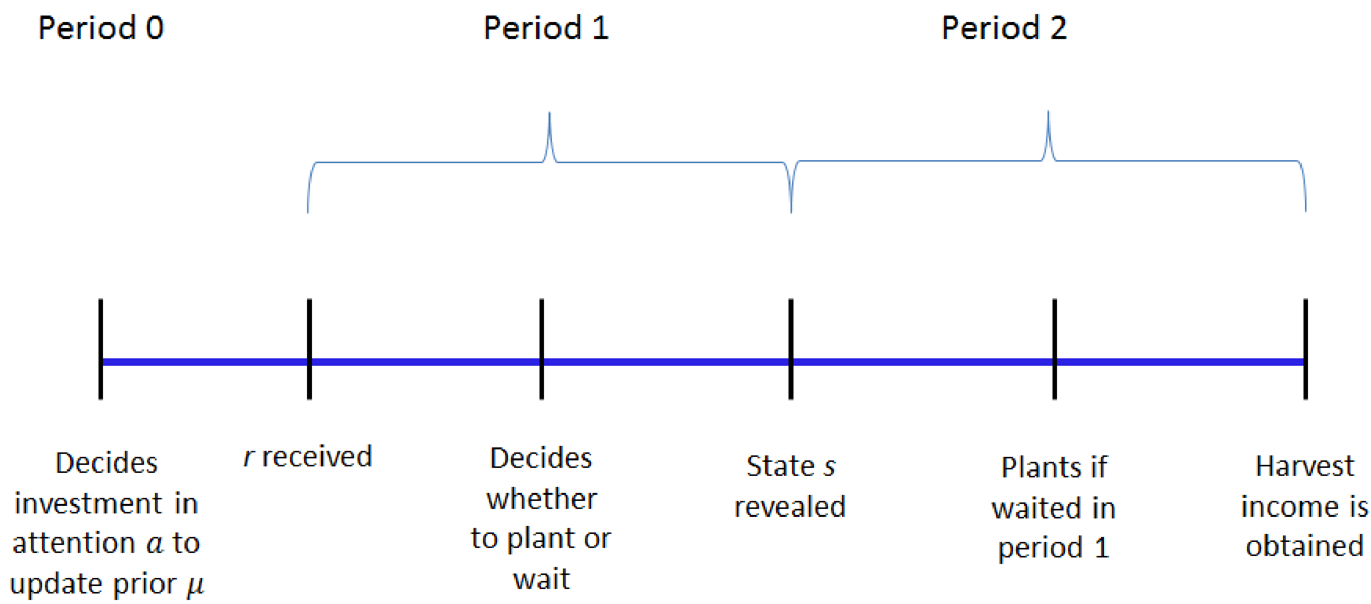


Figure 2. Optimal Planting Decision

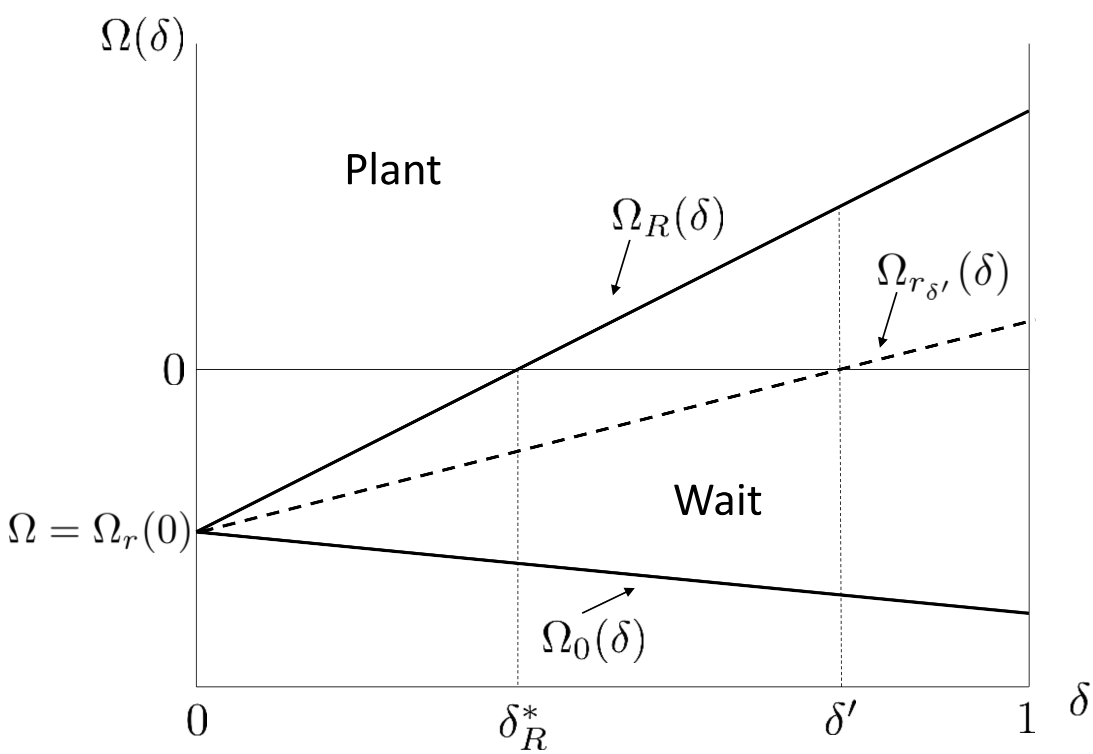


Figure 3. District Location in Andhra Pradesh, India

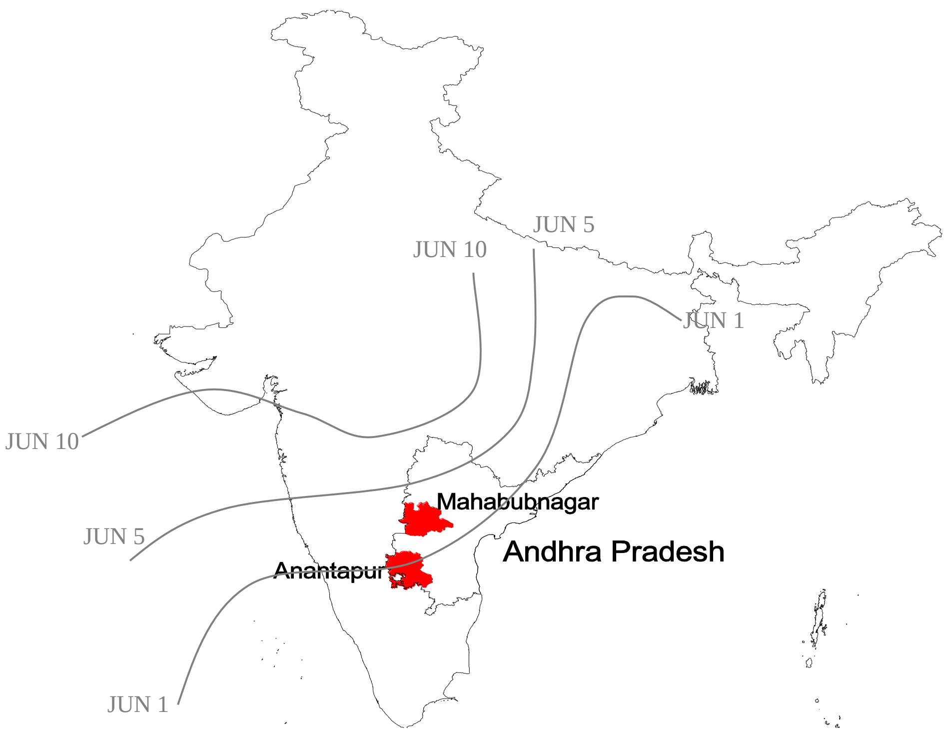


Figure 4 Onset of the Monsoon over Time

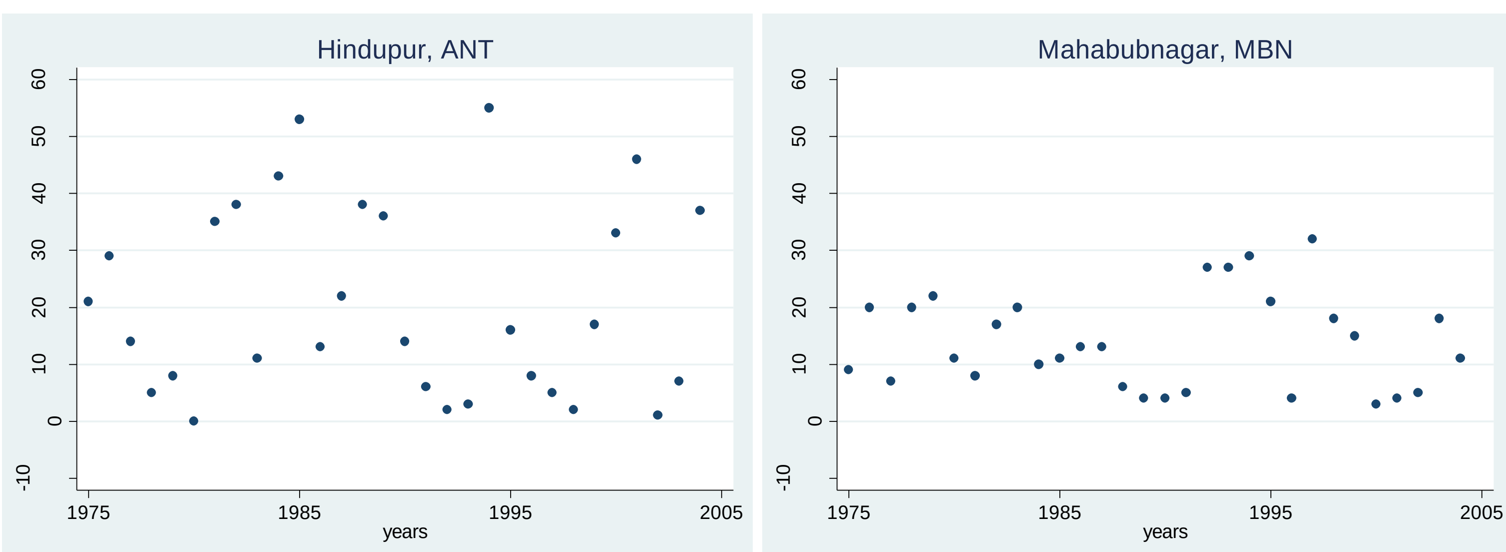


Figure 5. Days until Onset of the Monsoon since IMD Forecast over Time

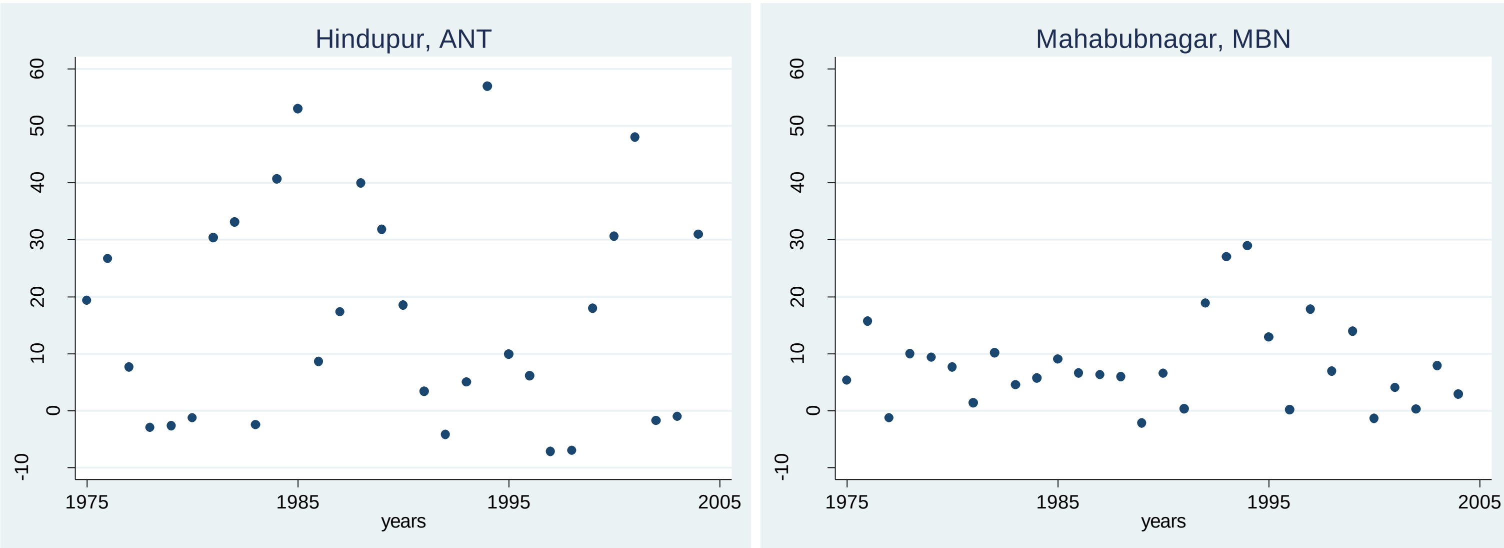


Table 1. Assignment of villages to rainfall stations in each district

Rainfall gauge	N. Villages	N. Households	Distance (miles)	N. Years of rainfall data
Mahabubnagar				
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**
Anantapur				
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: This table reports the number of villages assigned to each rainfall station and the number of years with available rainfall data. 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 rainfall 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. Summary Statistics

	Full Sample				Mahbubnagar	Anantapur	p-value
	Mean	Std. Dev.	Min	Max	Means		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Preferences							
Risk aversion	0.37	0.48	0.00	1.00	0.38	0.36	0.756
Discount Rate	0.30	0.28	0.00	2.33	0.26	0.37	0.001
Information / Social Networks							
Membership in BUA (1=Yes)	0.03	0.16	0.00	1.00	0.04	0.00	0.083
Weather info from informal sources (1=Yes)	0.19	0.39	0.00	1.00	0.18	0.22	0.337
Wealth							
Log market value of the house	10.85	0.77	8.70	13.12	10.93	10.69	0.034
Log value of owned land	11.87	1.00	7.15	16.86	12.12	11.40	0.000
Ability to smooth income shocks							
Participation in chit fund (1=Yes)	0.24	0.43	0.00	1.00	0.19	0.33	0.188
Household is credit constrained (1=Yes)	0.80	0.40	0.00	1.00	0.80	0.80	0.924
Log per capita total income	7.79	0.97	5.08	13.34	7.65	8.26	0.002
Exposure to Rainfall Shocks							
Cultivated land in acres	6.28	6.03	0.00	82.00	6.93	5.06	0.003
Pct. land of rainfed crops	0.73	0.29	0.00	1.00	0.70	0.78	0.030
Other Household Characteristics							
Literacy (1=Yes)	0.37	0.48	0.00	1.00	0.33	0.46	0.027
Education	3.31	4.36	0.00	18.00	2.91	4.04	0.036
Age of the household's head	47.83	11.91	21.00	80.00	47.22	46.08	0.131
Age of the eldest household member > 60 (1=Yes)	0.30	0.46	0.00	1.00	0.31	0.28	0.572
Forward caste (1=Yes)	0.15	0.36	0.00	1.00	0.12	0.21	0.137
Consumption (Rs.)	36,358	21,492	5,980	1,277,640	37,168	34,855	0.250
Land Characteristics							
Herfindahl index soil types	0.86	0.21	0.33	1.00	0.88	0.82	0.036
Pct. of land with loamy sand soil	0.50	0.45	0.00	1.00	0.51	0.48	0.680
Pct. of land with loam soil	0.21	0.38	0.00	1.00	0.16	0.31	0.064
Pct. of land with clay loam	0.19	0.35	0.00	1.00	0.22	0.13	0.257
Pct.of land with slope (higher than 1%)	0.45	0.46	0.00	1.00	0.48	0.41	0.081
Pct. of land with soil deeper than 80 cm	0.12	0.27	0.00	1.00	0.13	0.09	0.303
Onset of Monsoon distribution							
Ind. Def. of monsoon (mms.)	29.37	10.08	7.92	101.24	27.93	32.07	0.001
Mean of subjective distribution of monsoon onset	13.82	0.96	0.00	22.20	13.47	14.49	0.000
Std Dev of subjective distribution of monsoon onset	10.37	1.14	0.00	18.74	10.19	10.69	0.007
Mean of Actual Onset of monsoon (District definition)	13.97	1.03	12.86	16.29	13.93	14.05	0.477
Mean of Actual onset of monsoon (Individual defination)	13.99	1.08	12.57	16.64	13.96	14.06	0.553
Unconditional St. Dev. of actual onset of monsoon (Individual definition)	14.56	5.70	5.60	30.87	11.15	20.94	0.000
Conditional St. Dev. of actual onset of monsoon (Individual definition)	14.56	5.69	5.60	30.87	11.16	20.94	0.000
Agricultural Practices							
Attended meeting in 2004	0.07	0.26	0.00	1.00	0.08	0.05	0.172
Bought rainfall insurance in 2004	0.03	0.16	0.00	1.00	0.03	0.02	0.651
Bought insurance in 2004 conditional on attending meeting	0.37	0.48	0.00	1.00	0.34	0.48	0.155
Pct. expenses before monsoon	0.26	0.23	0.00	1.00	0.26	0.26	0.695
Average planting kartis	14.10	1.45	10.00	25.00	13.35	15.11	0.000
Household ever replanted (1=Yes)	0.22	0.41	0.00	1.00	0.30	0.07	0.000
Intercropping (1=Yes)	0.47	0.50	0.00	1.00	0.25	0.88	0.000
Agricultural Profits (Rs.)	-1,508	4,672	-67,455	17,135	-956	-2,559	0.001
Distance from village to rainfall gauge	4.36	1.75	0.62	8.29	4.08	4.87	0.168
Knowledge of good farmers							
Individual is considered a progressive farmer	0.30	0.46	0.00	1.00	0.28	0.33	0.205
Individual is considered a good crop producter	0.03	0.33	0.00	20.00	0.03	0.02	0.525
Individual correctly identifies good crop producers	0.02	0.12	0.00	1.00	0.02	0.02	0.933
Knowledge of abstract insurance (index)	0.32	0.42	0.00	1.00	0.31	0.34	0.533
Number of Observations							
2004	1050				687	363	
2006	941				629	312	
2008	1050				687	363	
2009	993				648	345	

Notes: This table reports the summary statistics of all the variables used in the analysis. The statistics reported are computed by weighting observations by the appropriate stratification (Attending Marketing Meeting and Purchasing Insurance in 2004) weights. AppendixTable A1 contains a detailed description of the variables reported. Unless otherwise noted, all variables were collected in 2004. Consumption, Actual onset of the monsoon and the percentage of expenses before the onset were collected in 2004, 2006 and 2009. The average planting kartis was collected in 2004, 2006, 2008 and 2009. For variables collected in multiple years, the statistics reported use all available years . The variable bought rainfall insurance is conditional on having attended a marketing meeting. Under "Number of observations" we report the number of observations in each year of data collection.

Table 3: Comparison of Subjective and Historical Distributions

District:	All			Mahbubnagar			Anantapur		
Distribution:	Subjective	Historical		Subjective	Historical		Subjective	Historical	
		Individual	District		Individual	District		Individual	District
					Means				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Support of Distributions									
Highest bin	15.0	15.3	15.5	15.0	15.3	14.8	15.6	17.7	16.9
Lowest bin	12.3	12.8	12.0	12.3	12.8	12.0	13.3	12.4	12.0
Number of bins with positive mass	3.5	4.2	4.3	3.7	3.6	3.8	3.2	5.5	5.2
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	0.9	1.0	1.0	0.9	0.8	0.9	0.8	1.4	1.3
Unimodal (1=Yes)	93.1%	70.4%	77.8%	92.1%	96.2%	93.1%	95.1%	21.9%	49.1%
Distributions is Normal (1=Yes)	95.5%	75.8%	89.3%	98.8%	82.8%	100.0%	89.3%	62.8%	69.0%
Accuracy									
Subjective distribution has same mode as historical distribution (1=Yes)		0.43	0.46	-	0.51	0.57	-	0.30	0.26

Notes: This table reports statistics about the support and moments of the subjective and historical distributions. The historical distribution is computed using the individual definition of monsoon or the district definition which is the average across individuals in the district. The table also reports measures of similaritybetween the subjective and historical distributions. The support of the distribution is measured in kartis or 10-day periods from April to September (kartis 9 to 20). Appendix Table A2 reports the name and dates of each kartis. "Number of bins with positive mass" corresponds to the number of kartis in the support with positive probability. The distribution is normal if one cannot reject that the kurtosis and skewness come from a normal distribution. The comparison of subjective and historical distributions is based on the chi-square statistic described in Section 3.3 and used in Tables 6-8.

Table 4. Behavior under uncertainty

Dependent variable	Rainfall required to plant (mms.)		Pct. expenses before monsoon	
Definition of Onset:	Individual	District	Individual	District
	(1)	(2)	(3)	(4)
Variable of interest				
St. Dev. of conditional historical distribution of monsoon onset	0.638***	0.553***	-0.003***	-0.002
	(0.098)	(0.076)	(0.001)	(0.001)
Preferences				
Risk aversion	0.549	0.457	0.009	0.009
	(0.591)	(0.589)	(0.010)	(0.011)
Discount rate (*100)	0.890	0.386	-0.012	-0.011
	(1.015)	(1.071)	(0.016)	(0.017)
Land characterstics				
Pct. of land with loamy sand soil	-2.629	-3.179	-0.006	-0.005
	(2.992)	(3.108)	(0.036)	(0.035)
Pct. of land with loam soil	7.631**	8.439**	-0.004	-0.010
	(3.282)	(3.412)	(0.036)	(0.035)
Pct. of land with clay loam	15.138***	16.460***	-0.015	-0.021
	(2.494)	(2.554)	(0.033)	(0.032)
Pct.of land with slope (higher than 1%)	-1.404**	-1.207*	-0.006	-0.008
	(0.648)	(0.625)	(0.010)	(0.010)
Pct. of land with soil deeper than 80 cm	8.942***	9.867***	0.002	-0.003
	(2.274)	(2.375)	(0.032)	(0.031)
Pct. land of rainfed crops	2.204**	1.270	-0.009	-0.006
	(1.053)	(1.094)	(0.016)	(0.016)
Household characteristics				
Forward caste (1=Yes)	1.773*	1.196	-0.017	-0.016
	(0.887)	(0.808)	(0.010)	(0.011)
Education	-0.014	-0.045	0.001	0.001
	(0.063)	(0.058)	(0.001)	(0.001)
Age of the household's head	0.045**	0.035*	0.000	0.000
	(0.021)	(0.020)	(0.000)	(0.000)
Cultivated land in acres	0.062	0.063	-0.002***	-0.002***
	(0.040)	(0.040)	(0.000)	(0.000)
Mean dependent variable	30.356	30.356	0.248	0.248
Observations	1046	1046	2,048	2,048
Individuals	1046		1046	
Years	2004		2004 and 2009	
R-squared	0.546	0.477	0.012	0.007

Notes: This table tests Proposition 1 of the model. In columns (1)-(3) the dependent variable is the amount of rain in millimeters that respondents require to start planting. In columns (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 the conditional historical distribution of the monsoon onset is the standard deviation of the number of days since 1st June until the onset of monsoon (using the relevant definition) minus the IMD forecast of the onset of the monsoon for that year, across all available years. Columns (1)-(2) and (4)-(5) use the individual definition of onset while columns (3) and (6) use the average of individual definitions (district definition). All regressions use OLS methods and 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 5. Do beliefs affect behavior?

	Bought insurance conditional on attending meeting	Average planting kartis	Household ever replanted	Pct. expenses before onset of the monsoon
	(1)	(2)	(3)	(4)
Variables of interest				
Subjective mean of onset	0.039***	0.095***	-0.053***	-0.000
	(0.010)	(0.029)	(0.019)	(0.005)
Ind. def of monsoon (mms.)	-0.000	0.001	-0.001	0.004***
	(0.001)	(0.002)	(0.003)	(0.000)
Actual onset of monsoon	-0.033*	0.104***	-0.029	-0.153***
	(0.017)	(0.022)	(0.132)	(0.014)
Preferences				
Risk aversion	-0.083***	-0.030	-0.012	0.013
	(0.013)	(0.042)	(0.029)	(0.008)
Discount Rate (*100)	-0.041*	-0.017	0.055	-0.015
	(0.023)	(0.099)	(0.050)	(0.015)
Land characteristics				
Pct. of land with rainfed crops	0.048**	-0.117	0.033	-0.026*
	(0.024)	(0.088)	(0.052)	(0.014)
Household characteristics				
Credit constrained (1=Yes)	-0.063***	-0.071	0.086***	0.002
	(0.015)	(0.051)	(0.031)	(0.009)
Village Fixed Effects	Yes	Yes	Yes	Yes
Other controls?	Yes	Yes	Yes	Yes
Mean dependent variable	0.44	14.12	0.25	0.25
Observations	5,800	3,293	931	2,016
Individuals	489	1032	931	1046
Years	2004	2004/06/08 and 2009	2006	2006 and 2009
R-squared	0.279	0.263	0.225	0.086

Notes: This table tests Proposition 2. Subjective mean of onset is the mean of the subjective distribution elicited from respondents. The regression in column (1) uses a stratification variable as the dependent variable and thus is run using Weighted Least Squares. Regressions in columns (2)-(4) control for variables used for stratification (attended marketing meeting and insurance purchase) and are run using OLS. In addition, all regressions control for Forward caste (1=Yes), Literacy (1=Yes), Age of the household's head, Market value of the house (expressed in Rs.100,000), Value of Owned land (expressed in Rs. 1,000,000), Use of intercropping (1=Yes), and village fixed effects. Standard errors are robust in columns 1-2 and 4 and clustered in columns 3 and 5. All standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively.

Table 6. Who has accurate beliefs?

Definition of Onset:	District			Individual		
			P-val of F-test			P-val of F-test
	(1)	(2)	(3)	(4)	(5)	(6)
Village characteristics						
St. Dev. of conditional historical distribution of monsoon onset	0.034***			-0.011	0.011	
	(0.010)			(0.009)	(0.021)	
Distance from village to rainfall gauge	0.069*			0.055**		
	(0.035)			(0.024)		
Preferences			0.433			0.513
Risk aversion	-0.069	-0.092		0.057	0.030	
	(0.062)	(0.071)		(0.041)	(0.040)	
Discount Rate	0.255**	0.037		0.090	0.069	
	(0.121)	(0.108)		(0.094)	(0.083)	
Wealth			0.082			0.007
Log (market value of the house/100,000)	0.124**	0.100**		0.107**	0.085**	
	(0.060)	(0.043)		(0.044)	(0.035)	
Log (value of owned land/1.000,000)	0.060	0.042		0.114*	0.097*	
	(0.069)	(0.052)		(0.060)	(0.050)	
Ability to smooth income shocks			0.390			0.005
Participation in chit fund (1=Yes)	0.348***	0.085		0.277***	0.151**	
	(0.101)	(0.072)		(0.077)	(0.069)	
Household is credit constrained (1=Yes)	-0.138*	-0.077		-0.146**	-0.139***	
	(0.077)	(0.060)		(0.059)	(0.043)	
Log (per capita total income/10,000)	0.167***	0.047		0.074**	0.002	
	(0.057)	(0.058)		(0.035)	(0.033)	
Exposure to Rainfall Shocks			0.073			0.004
Logarithm of cultivated land in acres	-0.265**	-0.038		-0.283***	-0.137**	
	(0.111)	(0.075)		(0.068)	(0.063)	
Pct. of land with rainfed crops	0.235*	-0.224*		0.035	-0.130*	
	(0.127)	(0.116)		(0.093)	(0.068)	
Information / Social Networks						
Weather info from informal sources (1=Yes)	-0.039	-0.005		-0.137	-0.100	
	(0.080)	(0.069)		(0.083)	(0.080)	
Other Household Characteristics			0.190			0.378
Herfindahl index soil type	0.133	0.196*		0.116	0.169*	
	(0.132)	(0.105)		(0.102)	(0.097)	
Literacy (1=Yes)	-0.022	-0.048		0.013	0.034	
	(0.073)	(0.060)		(0.056)	(0.049)	
Age of the household's head	-0.000	-0.003		0.003	0.001	
	(0.003)	(0.003)		(0.002)	(0.002)	
Age of the eldest household member > 60 (1=Yes)	-0.043	-0.058		-0.019	-0.035	
	(0.051)	(0.057)		(0.057)	(0.060)	
Forward caste (1=Yes)	0.117	-0.107		0.071	0.038	
	(0.104)	(0.092)		(0.065)	(0.057)	
Village Fixed Effects	No	Yes		No	Yes	
Mean dependent variable	1.53	1.53		1.99	1.99	
Observations	1024	1024		1024	1024	
R-squared	0.147	0.336		0.123	0.276	

Notes: This table tests Proposition 3. The dependent variable is the logarithm of the chi-square computed from comparing the subjective and the historical distributions. In colums 1-3 the historical distribution is computed using the district definition of onset of monsoon, that is the average in the district of the minimum quantum of rainfall that respondent requires to plant while columns 5-7 use individual definition of the onset of the monsoon. All regressions use OLS methods and control for the variables used in the stratification. Columns 3 and 7 include village fixed effects. Columns 4 and 8 report the p-value of an F-test that the coefficients of the variables under each heading are jointly zero. Robust standard errors are reported in brackets below the coefficient. The symbols *, ** and *** represent significance at the 10, 5 and 1 percent, respectively. See Appendix Table A1 for the definition of variables.

Table 7. Does Accuracy Matter?

Panel A: Yields and Profits

	Total yields		Total profits (in thousand Rs.)	
Definition of Onset:	Individual	District	Individual	District
	(1)	(2)	(3)	(4)
Log (χ^2) * No Paddy	-86.066***	-75.467***	125.893	-208.736
	(29.081)	(20.537)	(486.769)	(375.207)
Individual fixed effects	Yes	Yes	Yes	Yes
Crop fixed effects	Yes	Yes	Yes	Yes
Plot characteristics	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Mean of Dependent Variable	277.2	277.2	-1385.4	-1385.4
Number of Observations	3,678	3,678	2,128	2,128
Number of Individuals	994	994	963	963
Years	2004 and 2006		2004	
R squared	0.702	0.703	0.655	0.655

Panel B: Ever Replant and Consumption

	Household ever replanted				Consumption			
Definition of Onset:	Individual		District		Individual		District	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(χ^2)	0.055***	0.043**	0.046**	0.038**	-68.593	-452.987	-374.670	-769.088
	(0.021)	(0.019)	(0.019)	(0.019)	(480.471)	(536.718)	(508.252)	(524.067)
Subjective mean of onset	-0.048***	-0.038**	-0.074***	-0.058***	-585.513	-538.692	-410.529	-184.495
	(0.019)	(0.018)	(0.020)	(0.020)	(553.082)	(532.176)	(564.564)	(577.386)
Ind. def of monsoon (mms.)	-0.002	-0.001	-0.002	-0.002	-16.890	0.354	-55.664	-20.739
	(0.003)	(0.002)	(0.001)	(0.001)	(45.900)	(38.580)	(49.264)	(36.863)
Actual onset of monsoon	0.000	-0.040		-0.029	-1890.299***	-1130.817***	-2144.677***	-1316.806***
	(0.130)	(0.052)		(0.046)	(404.571)	(579.364)	(579.364)	(401.609)
Village fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plot characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	0.251	0.251	0.251	0.251	38509.07	38509.07	38509.07	38509.07
Observations	930	930	930	930	3003	3003	3003	3003
Individuals	930	930	930	930	1027	1027	1027	1027
Years		2006				2004,2006 and 2009		
R-squared	0.230	0.150	0.231	0.15	0.127	0.113	0.128	0.113

Notes: This table tests Proposition 4. In the odd columns of Panel A the log(χ2) is computed with the historical distribution that uses the district definition, that is average in the district of the minimum quantum of rainfall that respondent requires to plant while the even columns use the individual definition of the onset of the monsoon. In Panel A, the regression includes crop dummies, 2006 dummy, No paddy dummy x 2006 dummy, 2006 dummy x log(χ2) 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 80 cm (1=Yes) and use of intercropping (1=Yes). Each observation is a crop and if a household grows the same crop in more than one plot, then plot characteristics are weighted at the crop level by the area of the plot. The specifications in Panel A are estimated with OLS including household fixed effects and clustering standard errors at the household level. In Panel B, "Actual Start of Monsoon" in odd 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 column, the "district" definition, rather than each individual definition is used. In Panel B, 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, percentage of land with rain fed crops, credit constraints (1=Yes), forward caste (1=Yes), literacy (1=Yes), Age of the household's head, log market value of the house, log value of owned land and Use of intercropping (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.

Table 8: Knowledge

	Knowledge of abstract insurance (index)		Individual is considered a progressive farmer		Individual is considered a good crop producer		Individual correctly identifies good crop producers		Average planting Kartis	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log(χ^2)	-0.006	-0.021	-0.017	-0.031***	-0.039*	-0.039*	-0.009	-0.014*	0.061**	0.242***
	(0.015)	(0.014)	(0.013)	(0.010)	(0.021)	(0.021)	(0.007)	(0.008)	(0.024)	(0.047)
Village fixed effects	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable	0.492	0.492	0.343	0.343	0.142	0.142	0.047	0.047	14.1	14.1
Observations	1031	1031	2006	2006	1031	1031	1031	1031	3284	3284
Individuals	1031	1031	1050	1050	1031	1031	1031	1031	1032	1032
Years	2004		2004 and 2009		2004		2004		2004/06/08 and 2009	
R-squared	0.370	0.328	0.161	0.122	0.050	0.044	0.217	0.084	0.261	0.133

Notes: The regressions control for variables used in the stratification (attended marketing meeting and insurance purchase) and for the following additional controls: Risk aversion, Discount Rate, percentage of land with rainfed crops, credit constraints (1=Yes), forward caste (1=Yes), literacy (1=Yes), Age of the household's head, log market value of the house, log value of owned land, Use of intercropping (1=Yes), and village fixed effects. See Appendix Table A1 for a description of the variables. Standard errors are clustered in columns (3) and (4) and are robust in the rest of columns. All standard errors are reported in brackets below the coefficient. The symbols * and *** represent significance at the 10 and 1 percent, respectively.

Online Appendix

Proof of Propositions

This section provides the proofs to the propositions and discusses the case when $\widehat{\mu} > \mu^*$.

Proof of Proposition 1

Proof. First note that the assumptions about $f(r, \delta|s)$ ensure that the function $\Omega_r(\delta)$ is monotonic for all signals $r \in [0, R]$. In addition, $\Omega_r(0) < 0$ because $\widehat{\mu} < \mu^*$ and by definition $\Omega_{r_\delta}(\delta) = 0$, $r \in [0, R]$. Therefore $\frac{\partial \Omega_{r_\delta}(\delta)}{\partial \delta} > 0$ because $\Omega_{r_\delta}(\delta)$ is increasing in δ. Thus, $\frac{\partial \phi(r_\delta,\delta)}{\partial \delta} < 0$. Next, given the definition of $\Omega_{r_\delta}(\delta)$ in equation (2) in the text we can write

$$\phi(r_\delta, \delta) = \frac{\widehat{\mu}(\overline{y} - \underline{y})}{\underline{y}(1 - \widehat{\mu})}.$$

Then, by implicit differentiation we have

$$\frac{\partial r_\delta}{\partial \delta} = -\frac{\frac{\partial \phi(r_\delta,\delta)}{\partial \delta}}{\frac{\partial \phi(r_\delta,\delta)}{\partial r}} < 0,$$

because both $\frac{\partial \phi(r_\delta,\delta)}{\partial \delta} < 0$ and $\frac{\partial \phi(r_\delta,\delta)}{\partial r} < 0$. ■

Proof of Proposition 2

Proof. This follows easily from the fact that $\frac{\partial \Omega_r(\widehat{\mu})}{\partial \mu} > 0$ and that the farmer plants whenever $\Omega_r(\widehat{\mu}) > 0$. ■

Proof of Proposition 3

Proof. We first show that inaccurate priors, that is, the fact that the prior $\widehat{\mu}$ may differ from the objective probability μ, results in a loss in expected profits from rainfed crops. More formally,

$$EY_a(\mu) - EY_a(\widehat{\mu}) > 0.$$

When $\widehat{\mu} < \mu^*$, then either $\widehat{\mu} < \mu$ or $\widehat{\mu} \geq \mu$ because by assumption $\mu < \mu^*$.

If $\widehat{\mu} < \mu$, then $\delta^*_R(\mu) < \delta^*_R(\widehat{\mu})$. As a result,

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \int_{r_\delta}^{R} \pi_r \Omega_r(\mu)\, dr - \int_{\widehat{r}_\delta}^{R} \pi_r \Omega_r(\mu)\, dr.$$

Since $\widehat{\mu} < \mu$, we need to show that $\widehat{r}_\delta > r_\delta$. By definition, $\Omega_{\widehat{r}_\delta}(\widehat{\mu}, \delta) = \Omega_{r_\delta}(\mu, \delta) = 0$. Using the fact that $\widehat{\mu} < \mu$, this implies that $\phi(\widehat{r}_\delta, \delta) < \phi(r_\delta, \delta)$ and because $\phi(r, \delta)$ is decreasing in r, then $\widehat{r}_\delta > r_\delta$ as was to be shown. Therefore,

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \int_{r_\delta}^{\widehat{r}_\delta} \pi_r \Omega_r(\mu)\, dr,$$

and we can write the loss in expected profits as

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \begin{cases} 0 & \delta < \delta_R^*(\mu) \\ \int_{r_\delta}^{R} \pi_r \Omega_r(\mu)\, dr & \delta_R^*(\mu) \leq \delta < \delta_R^*(\widehat{\mu}) \\ \int_{r_\delta}^{\widehat{r}_\delta} \pi_r \Omega_r(\mu)\, dr & \delta \geq \delta_R^*(\widehat{\mu}) \end{cases} \tag{1}$$

The second line comes from the fact that a farmer with prior $\widehat{\mu}$ will wait but a farmer with accurate prior μ will plant if the signal satisfies $r > r_\delta$. The third line comes from the fact that because $r_\delta < \widehat{r}_\delta$, the farmer will plant if the signal $r \in [r_\delta, \widehat{r}_\delta]$ if she has priors μ but will wait if she has priors $\widehat{\mu}$. When the signal is $r > \widehat{r}_\delta$, the farmer will always plant. Both the second and third line of expression (1) are positive because $\Omega_r(\mu) > 0$ because $\delta > \delta_R^*(\mu)$ and $r > r_\delta$.

If on the other hand $\widehat{\mu} > \mu$, then $\delta_R^*(\mu) > \delta_R^*(\widehat{\mu})$ and $r_\delta > \widehat{r}_\delta$ using a similar argument as the one above. Therefore the loss in expected profits can be written as

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \begin{cases} 0 & \delta < \delta_R^*(\widehat{\mu}) \\ -\int_{\widehat{r}_\delta}^{R} \pi_r \Omega_r(\mu)\, dr & \delta_R^*(\widehat{\mu}) \leq \delta < \delta_R^*(\mu) \\ -\int_{\widehat{r}_\delta}^{r_\delta} \pi_r \Omega_r(\mu)\, dr & \delta \geq \delta_R^*(\mu) \end{cases} \tag{2}$$

The second line of expression (2) above is also positive because $\Omega_r(\mu) < 0$, for all signals $r \in [0, R]$ as $\delta < \delta_R^*(\mu)$, and the third line is also positive because $r < r_\delta$ and thus $\Omega_r(\mu) < 0$.

If we now consider the case $\widehat{\mu} \geq \mu^*$, the bound $\delta_R^*(\widehat{\mu})$ is replaced by $\delta_0^*(\widehat{\mu})$. As above, we have two cases two consider: $\delta_0^*(\widehat{\mu}) < \delta_R^*(\mu)$ and the opposite. If $\delta_0^*(\widehat{\mu}) < \delta_R^*(\mu)$ we obtain

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \begin{cases} \underline{y} - \mu\overline{y} & \delta < \delta_R^*(\widehat{\mu}) \\ -\int_{\widehat{r}_\delta}^{R} \pi_r \Omega_r(\mu)\, dr & \delta_0^*(\widehat{\mu}) \leq \delta < \delta_R^*(\mu) \\ -\int_{\widehat{r}_\delta}^{r_\delta} \pi_r \Omega_r(\mu)\, dr & \delta \geq \delta_R^*(\mu) \end{cases} \tag{3}$$

which is similar to expression (2) except for the bounds in the second line and the first line because a farmer with priors $\widehat{\mu}$ will now plant. Since by assumption $\mu < \mu^*$, $\underline{y} - \mu\overline{y} > 0$ and so $EY_a(\mu) - EY_a(\widehat{\mu}) > 0$.

On the other hand, if $\delta_0^*(\mu) > \delta_R^*(\mu)$, the loss in expected profits can be written as

$$EY_a(\mu) - EY_a(\widehat{\mu}) = \begin{cases} \underline{y} - \mu\overline{y} & \delta < \delta_R^*(\widehat{\mu}) \\ \underline{y} - \mu\overline{y} + \int_{r_\delta}^{R} \pi_r \Omega_r(\mu)\, dr & \delta_R^*(\mu) \leq \delta < \delta_0^*(\widehat{\mu}) \\ -\int_{\widehat{r}_\delta}^{r_\delta} \pi_r \Omega_r(\mu)\, dr & \delta \geq \delta_0^*(\widehat{\mu}) \end{cases} \tag{4}$$

where the second line is positive because $\underline{y} - \mu\overline{y} > 0$ and $\Omega_r(\mu) > 0$ when signal r satisfies $r > r_\delta$ and the third line is also positive because whenever $r < r_\delta$, $\Omega_r(\mu) < 0$.

As a result, expressions (1) - (4) are all non-negative and thus

$$EY_a(\mu) - EY_a(\widehat{\mu}) > 0 \quad \text{as was to be shown.}$$

Now, because $\nu = |\widehat{\mu} - \mu|$, if $\widehat{\mu} < \mu$, then

$$\frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \nu} \geq 0 \quad \text{if and only if} \quad \frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \widehat{\mu}} \leq 0.$$

Similarly, if $\widehat{\mu} > \mu$, then

$$\frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \nu} \geq 0 \quad \text{if and only if} \quad \frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \widehat{\mu}} \geq 0.$$

We therefore need to show that

$$\frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \widehat{\mu}} \leq 0 \quad (\geq 0) \quad \text{if} \quad \widehat{\mu} < \mu \quad (\widehat{\mu} > \mu).$$

First, one can show that $\frac{\partial \delta_R^*(\widehat{\mu})}{\partial \widehat{\mu}} < 0$ by implicit differentiation of $\Omega_R(\widehat{\mu}, \delta_R^*) = 0$. Similarly, one can show that $\frac{\partial \widehat{r}_\delta}{\partial \widehat{\mu}} < 0$ by implicit differentiation of $\Omega_{\widehat{r}_\delta}(\widehat{\mu}, \delta) = 0$.

If $\widehat{\mu} < \mu$, the loss in expected profits is given by expression (1). Thus, differentiating under the integral,

$$\frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \widehat{\mu}} = \begin{cases} 0 & \delta < \delta_R^*(\widehat{\mu}) \\ -\frac{\partial \widehat{r}_\delta}{\partial \widehat{\mu}} \pi_{\widehat{r}_\delta} \Omega_{\widehat{r}_\delta}(\mu) < 0 & \delta > \delta_R^*(\widehat{\mu}), \end{cases} \tag{5}$$

where the second line is negative because $\frac{\partial \widehat{r}_\delta}{\partial \widehat{\mu}} < 0$ as was shown and $\Omega_{\widehat{r}_\delta}(\mu) < 0$ because $\widehat{r}_\delta < r_\delta$.

If $\widehat{\mu} > \mu$, the loss in expected profits is given in (2) if $\widehat{\mu} < \mu^*$ and in (3)-(4) if $\widehat{\mu} > \mu^*$. Again differentiating under the integral, we obtain a very similar expression

$$\frac{\partial\left[EY_a(\mu) - EY_a(\widehat{\mu})\right]}{\partial \widehat{\mu}} = \begin{cases} 0 & \delta < \delta_i^* \\ \frac{\partial \widehat{r}_\delta}{\partial \widehat{\mu}} \pi_{\widehat{r}_\delta} \Omega_{\widehat{r}_\delta}(\mu) > 0 & \delta > \delta_i^* \end{cases} \tag{6}$$

where $\delta_i^* = \delta_R^*(\mu)$ for $\widehat{\mu} < \mu^*$ and $\delta_i^* = \delta_0^*(\widehat{\mu})$ for $\widehat{\mu} > \mu^*$. The only difference is the negative sign in the second line of expression (5) that does not appear in (6). Therefore, putting expressions (5) and (6) together,

$$\frac{\partial \left[EY_a(\mu) - EY_a(\widehat{\mu}) \right]}{\partial \nu} \geq 0,$$

as was to be shown. ■

Proof of Proposition 4

Proof. The proof that $\frac{\partial \Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial y_0} < 0$ is trivial. Now, $\frac{\partial \Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial \delta} = 0$ if $\widehat{\mu} < \mu^*$ and $\delta < \delta_R^*(\widehat{\mu})$ or $\widehat{\mu} > \mu^*$ and $\delta < \delta_0^*(\widehat{\mu})$. However, if $\widehat{\mu} < \mu^*$ and $\delta > \delta_R^*(\widehat{\mu})$ or $\widehat{\mu} > \mu^*$ and $\delta > \delta_0^*(\widehat{\mu})$, then differentiating under the integral

$$\frac{\partial \Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial \delta} = \int_{\widehat{r}_\delta}^{R} \left[\frac{\partial \widehat{\pi}_r}{\partial \delta} \Omega_r(\widehat{\mu}) + \widehat{\pi}_r \frac{\partial \Omega_r(\widehat{\mu})}{\partial \delta} \right] dr = \int_{\widehat{r}_\delta}^{R} I \, dr$$

where I can be simplified to

$$I = \widehat{\mu}(\overline{y} - \underline{y}) \frac{f(r, \delta | M)}{\delta} - (1 - \widehat{\mu}) \underline{y} \frac{f(r, \delta | D)}{\delta} > 0,$$

which is positive given the assumptions about $f(r, \delta | s)$. Therefore,

$$\frac{\partial \Pr\left[\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) > y_0\right]}{\partial \delta} \geq 0,$$

as was to be shown. ■

Case II: $\widehat{\mu} \geq \mu^*$

The paper focuses on the case where $\widehat{\mu} < \mu^*$ (Case I) and in this section we provide the analogous expressions for the case when the prior satisfies $\widehat{\mu} \geq \mu^*$. Since $\widehat{\mu} \geq \mu^*$, we have that $\Omega_r(0) = \Omega \geq 0$, for all signals r. In contrast to Case I in the text, a farmer that decides not to update in period 0 will always plant in period 1.

It is still the case that $\Omega_R(\delta)$ ($\Omega_0(\delta)$) is increasing (decreasing) in δ, but the level of informativeness δ_R^* defined before such that $\Omega_R(\delta_R^*) = 0$ no longer exists for $\delta \in [0, 1]$.[1] In

[1] The function $\Omega_R(\delta)$ is increasing and both values in the endpoints of the interval are positive, $\Omega_R(0) > 0$ and $\Omega_R(1) > 0$. No root therefore exists in the interval $\delta \in [0, 1]$.

contrast, there exists a level of informativeness δ_0^* such that $\Omega_0(\delta_0^*) = 0$. It is easy to show that $\delta_0^* < 1$ as well.[2] Using an argument analogous to the one made for Case I in the text, if informativeness satisfies $\delta < \delta_0^*$, then it is always optimal to plant in period 1, regardless of the signal r received. If however $\delta > \delta_0^*$, then the same cutoff r_δ that satisfies $\Omega_{r_\delta}(\delta) = 0$ exists. The farmer will wait if $r < r_\delta$ but will plant otherwise.

Appendix Figure 1 plots the function $\Omega_r(\delta)$ for the case when $\widehat{\mu} > \mu^*$ and thus $\Omega = \Omega_r(0) > 0$ for $r = R, r = 0$ and $r = r_{\delta'}$. The optimal planting strategy for the farmer with $\delta = \delta' \geq \delta_0^*$, is to plant if signal $r \in [r_\delta, R]$ and to wait if signal $r \in [0, r_\delta)$. In contrast, if $\delta < \delta_0^*$, the optimal planting strategy is to plant regardless of signal r.

Expected Profits

If $\widehat{\mu} \geq \mu^*$ (Case II) then expected profits $\widehat{E}Y_a$ when the prior is updated can be written as

$$\widehat{E}Y_a(\widehat{\mu}) = \begin{cases} \widehat{\mu}\overline{y} & \delta < \delta_0^*(\widehat{\mu}) \\ \underline{y} + \int_{\widehat{r}_\delta}^{R} \widehat{\pi}_r \Omega_r(\widehat{\mu})\, dr & \delta \geq \delta_0^*(\widehat{\mu}). \end{cases} \tag{7}$$

Expected profits in expression (3) of the text and in (7) above differ in two respects. First, the relevant informativeness threshold is different (δ_R^* in the text and δ_0^* here) and second, the fact that in expression (3) of the text the farmer waits when informativeness is below the threshold, while under expression (7) the farmer plants.

Choice of Attention

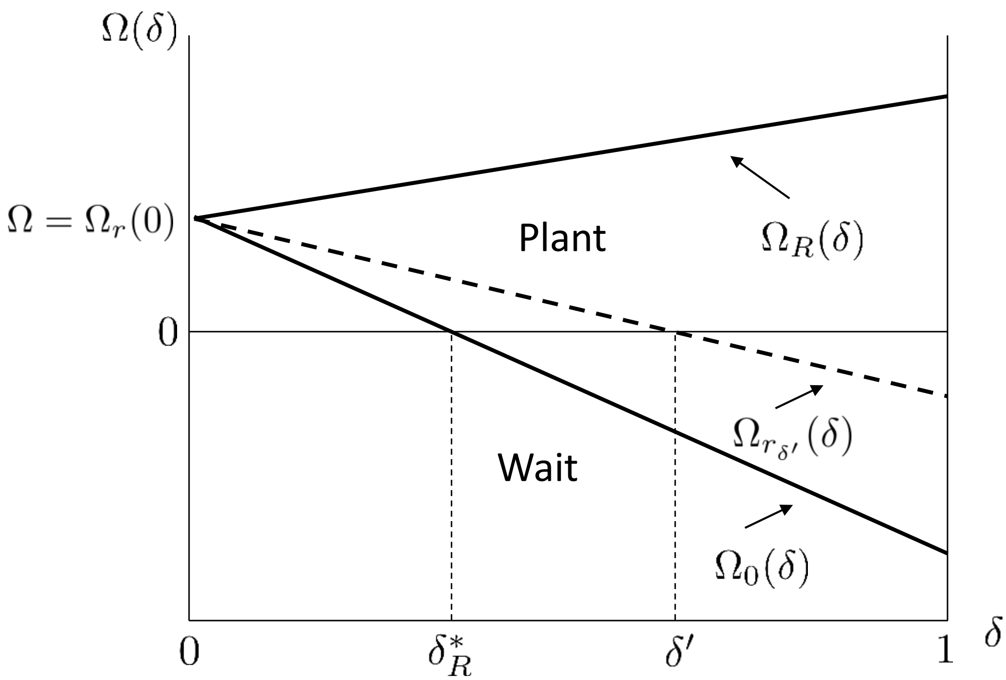
From expressions (3) in the text and (7) above we have

$$\widehat{E}Y_a(\widehat{\mu}) - \widehat{E}Y_0(\widehat{\mu}) = \begin{cases} 0 & \delta < \delta_0^*(\widehat{\mu}) \\ \underline{y} - \widehat{\mu}\overline{y} + \int_{\widehat{r}_\delta}^{R} \widehat{\pi}_r \Omega_r(\widehat{\mu})\, dr & \delta \geq \delta_0^*(\widehat{\mu}) \end{cases} \tag{8}$$

It is also clear from expression (8) that the decision to update the prior will depend on the returns to the alternative activity y_0 and informativeness δ.

[2] By assumption $f(0, 1|M) = 0$, therefore $\Omega_0(1) = -\underline{y} < 0$. Since the prior satisfies $\widehat{\mu} > \mu^*$, the difference in expected utility when $r = 0$ and $\delta = 0$, $\Omega_0(0) > 0$. By Bolzano's theorem, the continuous function $\Omega_0(\delta)$ with values of opposite signs in the interval $\delta \in [0, 1]$ has a root, and therefore δ_0^* exists. By continuity, since $\Omega_0(1) < 0$ and $\Omega_0(\delta_0^*) = 0$, $\delta_0^* < 1$.

Appendix Figure 1. Optimal Planting Decision under Case II ($\hat{\mu} > \mu^{*}$)



Appendix Table A1. Construction of Variables

Variable name	Significance of variable
Risk aversion	Dummy variable equal to 1 safe bet (50 Rs. if heads or tails) is chosen.
Discount Rate	(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)	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)	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)	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)	Present market value of the owned plots
Participation in chit fund (1=Yes)	Dummy variable equal to 1 if household participates in at least one chit fund (ROSCA)
Household is credit constrained (1=Yes)	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)	Total income per household member
Cultivated land in acres	Total land cultivated by the household
Pct. of land with paddy	Land dedicated to paddy over total household cultivated land
Pct. land of rainfed crops	Pct of total cultivated acres for the following crops: Sorghum, Groundnut, Castor, Red Gram, Maize, Cotton, All Millets, Sesamamum, Safflower and Greengram.
Literacy (1=Yes)	Dummy variable equal to 1 if the household head is literate
Education	Years of education of household's head
Age of the household's head	Age of household's head
Forward caste (1=Yes)	Dummy variable equal to 1 if the household is of forward caste
Pct. of land with loamy sand soil	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	Pct of total acres cultivated by the HH that have loam soil (red loam soil)
Pct. of land with clay loam	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%)	Pct of total acres cultivated by the household that have slope greater than 1%
Pct. of land with soil deeper than 80 cm	Pct of total acres cultivated by the household that are deeper than 80cm.
Pct. of land with rain sensitive crops	Pct of land under maize, red gram, green gram, black gram, finger millet, foxtail millet, sesamum, sorghum
Individual definition of start of monsoon (mms)	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	Percentage of total inputs' expenditure invested before the onset of the monsoon.
Dummy Anantapur	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	Distance from village to rainfall gauge in kms.
Age of the eldest household member > 60 (1=Yes)	1 if the eldest member of the household is more than 60 years old.
Attended meeting and bought insurance	Dummy variable equal to 1 if the household attended the weather insurance meeting and bought insurance
Bought insurance	Dummy variable equal to 1 if the household bought rainfall insurance
Average planting kartis	Average kartis when planting took place across crops in relevant year
HH replanted in last 10 years (1=Yes)	Dummy variable equal to 1 if the household has ever replanted at least one crop
Individual is considered a progressive farmer	Dummy variable equal 1 if an individual is considered a progressive farmer
Individual is considered a good crop producter	Number of times an individual is reported as best rainfed crop producer by other respondents.
Individual correctly identifies good crop producers	Dummy variable equal 1 if respondent reports an individual as best rainfed crop producer whose yields are in the top 25th percentile of yields in the village.
Knowledge of abstract insurance (index)	The index is composed of five questions related to the rainfall insurance. A correct answer to the question is coded as 1 and 0 otherwise. The answer to each question is added and divided by 5 to create the index.
Subjective mean of onset	Mean kartis of the subjective distribution
Actual onset of monsoon	Actual onset of the monsoon in relevant year computed using either the individual definition or the average of the individual definitions in the district.

Appendix Table A2. 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.

Appendix Table A3. Summary Statistics

	Full Sample				Mahbubnagar	Anantapur	p-value
	Mean	Std. Dev.	Min	Max	Means		
Actual onset of monsoon (District definition)							
2004	15.03	1.12	13.36	16.29	14.97	15.14	0.593
2006	13.50	0.41	12.86	13.93	13.46	13.57	0.319
2009	13.38	0.25	13.00	13.64	13.36	13.43	0.300
Actual onset of monsoon (Individual definition)							
2004	15.04	1.14	13.14	16.64	14.99	15.13	0.653
2006	13.54	0.44	12.71	15.71	13.44	13.71	0.017
2009	13.42	0.61	12.57	16.00	13.44	13.39	0.695
Average planting kartis							
2004	14.14	0.98	11.50	25.00	13.77	14.83	0.000
2006	14.21	1.61	11.00	22.00	13.69	15.24	0.000
2008	13.88	1.17	11.00	16.00	13.27	14.91	0.000
2009	14.22	2.25	10.00	21.00	13.48	16.17	0.000
Pct. expenses before monsoon							
2004	0.25	0.21	0.00	0.97	0.35	0.05	0.000
2009	0.28	0.26	0.00	1.00	0.17	0.47	0.000