LEARNING, ADAPTATION, AND CLIMATE UNCERTAINTY: EVIDENCE FROM INDIAN AGRICULTURE

NAMRATA KALA†

ABSTRACT. The profitability of many agricultural decisions depends on farmers' predictions about the weather. Climate change implies (possibly unknown) changes in the weather distribution. I study how farmers form predictions about a weather-dependent decision, the planting time, using rainfall signals. To capture the potential uncertainty caused by climate change, I develop an empirical framework that estimates, and finds support for, a general robust learning model in which farmers believe that the rainfall signals are drawn from a set of rainfall distributions. This indicates that farmers respond to greater (Knightian) uncertainty by modifying their predictions to be robust to such uncertainty.

† MIT SLOAN SCHOOL OF MANAGEMENT, NBER, BREAD AND JPAL. KALA@MIT.EDU, NAMRATAKALA.COM
Date: December 2019.
I am grateful to Michael Boozer for his invaluable guidance on this project. I would like to thank Dean Karlan, Robert Mendelsohn, Tavneet Suri, and Chris Udry for their guidance and support, and Achyuta Adhvaryu, Abhijit Banerjee, Prashant Bharadwaj, Gharad Bryan, Christopher Costello, Rahul Deb, Eric Edmonds, Jose-Antonio Espin-Sanchez, James Fenske, Xavier Gine, Daniel Keniston, Dominic Leggett, Robert McMillan, Anant Nyshadham, Rohini Pande, Debraj Ray, and seminar participants at Arizona State University, Boston College, Boston University, the Harris School at the University of Chicago, NBER Summer Institute Development meeting, Northeast Workshop on Energy Policy and Environmental Economics at Dartmouth, Pennsylvania State University, Rice University, Stanford Graduate School of Business, Stanford Environmental and Energy Policy Analysis Center Research Workshop, Yale University, University of Pennsylvania, University of Rochester, University of San Diego, and University of Toronto for helpful comments. All remaining errors are my own.

"The real challenge of dealing effectively with climate change is recognizing the value of wise and timely decisions in a setting where complete knowledge is impossible"

Intergovernmental Panel for Climate Change (IPCC), 2014

1. Introduction

Both economists and psychologists have established that decision making differs in risky and uncertain environments i.e. in settings where the probabilities of events are known versus those where they are not.[1] Climate change is a particularly relevant example of an environment characterized by unknowable uncertainty (where the probabilities of weather events may not be known), especially in the short and medium-run because it will manifest itself as changes in the weather distribution (Ferro, Hannachi, and Stephenson, 2005). Weather outcomes provide noisy signals of the underlying climate distribution, and climate change might do more than add to the complexity (and non-stationarity) of the stochastic process driving the weather and climate: it might also create uncertainty wherein the underlying stochastic process is no longer known. Furthermore, the historical record under recent climate change might be insufficient to statistically distinguish one particular climate process from another. Unlearnable uncertainty might thus be an important factor that determines climate-sensitive decisions, and one that has largely been overlooked in previous studies on adaptation. Indeed, if agents facing climate change are cognizant of this uncertainty and adapt accordingly, ignoring this aspect of decision making may misstate the extent of adaptation.

Climate change is estimated to have already affected food security adversely, and is projected to do so to a greater extent in the future(IPCC, 2013). This paper studies how farmers in India, a country highly vulnerable to climate change (IPCC, 2013), dynamically make an agricultural decision that is an important determinant of profits (the choice of planting time), and whose profitability depends on the timing of monsoon rainfall.[2] The high degree of uncertainty governing the monsoon has been exacerbated by recent decadal shifts in parts of the monsoon distribution in some areas (Goswami, Kulkarni, Mujumdar, and Chattopadhyay, 2010). The extent to which farmers respond to this greater uncertainty by changing the way they form weather predictions and their consequent behavior is the focus of this paper.

To account for the uncertainty introduced by climate change, I model learning by assuming that farmers believe that the rainfall signals they observe are drawn from an unknown member of a set of (unspecified) stochastic processes near an approximating model (following Hansen and Sargent (2008)). Put differently, this generalizes a standard Bayesian learning environment (which as I discuss below is a special case of the model) by allowing for the fact that farmers may not know the true stochastic process governing the weather.

[1] The classical reference for this is Ellsberg (1961). Gilboa (2009) and Hansen and Sargent (2008) provide a thorough theoretical treatment of these and related topics.

[2] Precipitation patterns are a crucial determinant of returns in agriculture, and the monsoon season in India is particularly important for agricultural returns. The literature documenting the impact of monsoon rainfall on agricultural returns is extremely extensive and impossible to cover in complete detail - Giné, Townsend, and Vickery (2007); Jacoby and Skoufias (1998), Rosenzweig and Binswanger (1993), and (Taraz, 2017) cover important aspects of how monsoon rainfall impacts agricultural yields and incomes in India.

The behavioral implication of this model is that farmers make decisions that are robust to model misspecification (i.e. not being certain what the rainfall distribution is) by exhibiting a concern for worst-case profits.

I estimate the model and find support for its predictions in long-term panel data from India on monsoon realizations and household-level data on agricultural decisions and returns. The agricultural decision I study, the timing of planting, contains information about household expectations about the monsoon, and is an economically crucial decision for farmers: a every 1% deviation from the optimal planting time in a given year causes about 3% lower profits relative to mean profits. Specifically, I find that farmers indeed exhibit a concern for worst-case profits when evaluating the uncertain relationship between the rainfall signal and the optimal planting time. Furthermore, the estimates provide additional support for the model by showing that the belief that rainfall signals are drawn from a set, rather than a single distribution, is much more pronounced in villages that experienced changes in the monsoon onset distributions in the last five decades. As an additional model-free, data driven validation of the model, I show that the maximum (i.e. (worst-case)) forecast errors are also much lower in villages that experienced changes in the monsoon onset distributions. This indicates that farmers respond to greater (Knightian) uncertainty in their environment by modifying their predictions to be robust to such uncertainty (this is not an implication of the alternative canonical models I additionally consider). Methodologically, the empirical framework I develop is quite general (and not tied to my application per se) and can be employed to test across learning models in other environments with unlearnable uncertainty.

As mentioned above, my framework additionally allows me to quantify the importance of robust decision making in my application by contrasting the estimates with those from two other canonical Bayesian learning models. Unlike robust learning, in these latter models, the farmer is assumed to know the *exact* stochastic processes that generate both the optimal planting time and the rainfall signals. All three models generate distinct testable predictions on how the farmer combines the information from past rainfall signals to determine her optimal planting time. In particular, the models yield structural equations governing the learning process which can be estimated.

Specifically, I contrast a robust normal learning model (RNLM) with two Bayesian normal learning models (NLMs) in which the mean optimal planting time is assumed to be time-invariant and time-varying respectively. The time invariant Bayesian NLM (or NLM for short) is perhaps one of the most widely used learning models in both the theoretical and empirical literature. In this, both the parameter of interest that is being learned (the mean optimal planting time) and the signals that provide information (rainfall signals) are drawn from *known* normal distributions. Versions of this basic model have been employed in numerous empirical microeconomic studies on learning; for instance, in the work studying the returns to new technologies or to sectors (see, for example, Foster and Rosenzweig (1995), Gibbons, Katz, Lemieux, and Parent (2005), Conley and Udry (2010)). Qualitatively, one of the main implications of this model is that the posterior estimate of the parameter being learned is formed by taking a linear combination of past signals where each signal is given equal weight (since the parameter of interest that the agents are learning about is time-invariant, each signal contains the same amount of information).

By contrast, in the robust NLM (introduced by Hansen and Sargent (2008) in the macroeconomics literature), the farmer starts off with an *approximating* model that consists of known normal distributions from which the time-invariant mean optimal planting time and the rainfall signals are drawn. She chooses an optimal planting time to maximize the worst case profits that could arise if the *true* data generating process was not necessarily the approximating model but could instead be any one from a set of stochastic processes that are close to it (loosely speaking, the size of this set determines the degree of robustness). In this sense, decision making is *robust* to the fact that the approximating model may be misspecified. An attractive feature of the robust NLM is that the optimal planting time is still a linear function of past signals; however, the qualitative difference is that these signals are not weighted equally. Specifically, more weight is placed on recent signals. Intuitively, earlier signals affect decision making in more time periods and hence, are more likely to be misleading in the worst case (and should hence be underweighted).

Another natural reason that could explain the overweighting of recent signals is that the farmers believe the mean optimal planting time is time-varying: by definition, the most recent signal would then be the most informative. I explore this alternative by considering a NLM where mean optimal planting time evolves following a known Gaussian AR (1) process. Here too, the farmer's optimal choice of planting time is a linear function of past signals (in which recent signals have greater weight). However, the precise structural equations that govern this behavior differ from those of the robust NLM and this is what allows me to contrast the parameter estimates. It is worth emphasizing that each of these models is economically interesting in their own right and the one that best explains the data depends on the context and is, ultimately, an empirical question. Hence, the empirical framework can be viewed as a methodological contribution in that it provides a parsimonious way to test across canonical learning models that embody varying beliefs about risk and uncertainty, and can be employed in other learning environments.

With this analysis, we can understand farmer decision making by testing across learning models, as well as test whether and how climate uncertainty impacts this behavior. This implies that in environments characterized by unlearnable uncertainty, robust learning may be an effective adaptation technique, and it is one that farmers are already employing to adapt to climate change, at least with regard to this one important agricultural decision.

This study makes two main contributions. Firstly, I introduce robust learning and decision making to the economics of climate change adaptation. It has been argued that changes in the weather affect agricultural returns (for a recent example, see Burke and Emerick (2013) who estimate the impacts of changes in decadal mean temperature and precipitation on crop yields) and that farmers adapt to these changes (for instance, Taraz (2017) shows that farmers respond to variation in mean rainfall by investing in irrigation). To assess the efficacy and relative importance of various policies that allocate adaptive resources, it is important to understand farmers' adaptation behavior. An important aspect of this adaptation behavior is using weather information to form beliefs about the changing climate(Lemoine, 2018). Ignoring unlearnable uncertainty, especially when analyzing weather sensitive behavior during climate change, can potentially be a crucial omission. These results also underscore the limitations of providing weather forecast information in absence of other adaptation measures, since weather forecast information may not allow farmers to discern which distribution the weather is drawn from under climate change, and so provides limited adaptive possibilities to the uncertainty caused by climate change. Finally, I show that farmers respond not only to changes in the mean weather, but also changes in the weather distribution more generally, and do so by changing the way they form weather predictions which impact their economic decisions.

Secondly, as mentioned above, a methodological contribution of this paper is to the literature that tests across learning models. In terms of its data requirements, such testing is challenging as long term panel data (on the signals observed and the decisions made in response) is needed. A consequence is that most micro-economic studies on testing learning have been conducted using data from the laboratory (Camerer and Ho, 1999; Cheung and Friedman, 1997) or from experimental games in the field (Barham, Chavas, Fitz, Ríos-Salas, and Schechter, 2014). To the best of my knowledge, this is the first paper to test across learning models while explicitly allowing for robust learning using individual-level field data on an important economic decision.

The rest of the paper is organized as follows. Section 1.1 discusses the related literature in more detail, while section 2 provides the background and context of the setting. Section 3 details the models of learning I test across, and section 4 provides the estimating equations that are implied by the models. Section 5 discusses the data used in estimation, 6 the results of the estimation, and 7 concludes.

1.1. *Related Literature*

Farmers make several decisions that incorporate information about their expectations of climate and weather, and which in turn affects their returns. Examples include irrigation (Kurukulasuriya, Kala, and Mendelsohn, 2011), crop choice (Kurukulasuriya and Mendelsohn, 2008; Miller, 2013) and crop varieties (Deressa, Hassan, Ringler, Alemu, and Yesuf, 2009). More recently, Lemoine (2018) contains a comprehensive discussion of the implications of inferring climate impacts from weather observations, especially in environments of irreversible investments. Learning about climate change for adaptation was first considered by Kelly, Kolstad, and Mitchell (2005), who model an agent who is learning about a climate which has changed to a new time-invariant state, and use agricultural data from Midwestern US states to estimate damages of adjusting to this new equilibrium climate. Taraz (2017) finds that farmers in India are less likely to invest in irrigation following a decade of higher than average rainfall, which is in keeping with a model of adaptation to decadal monsoon rainfall cycles. The choice of planting dates, the decision I study, is an important and relatively low-cost adaptation measure to perceived climate risk but there are fewer studies on planting dates as long term panel data of this decision is relatively rare.

There has also been considerable interest in studying how farmers' beliefs impact behavior. In recent important work, Giné, Townsend, and Vickery (2015) show that subjective beliefs regarding onset are not only a good approximation of actual onset, but also play an important role in planting decisions. Thus, they show that farmer beliefs predict their behavior, which is important for my estimation, since I do not have data on beliefs, and infer them from farmer choices (timing of planting decisions). Deressa, Hassan, Ringler, Alemu, and Yesuf (2009) find that information regarding weather and climate to be an important determinant in adoption of adaptation mechanisms for farmers in the Nile Basin in Ethiopia. The literature that examines the impact of forecasts on farmer beliefs, decisions and economic outcomes, finds mixed impacts depending on the context. Lybbert, Barrett, McPeak, and Luseno (2007) find that while herders in Kenya do update their subjective beliefs about future weather realizations in response to official forecasts, they do not respond to the forecast while taking decisions regarding their livestock, likely because their moving costs are low. Rosenzweig and Udry (2013) find that in areas where monsoon forecasts are reliable, they are an important determinant in planting investment decisions for Indian farmers.

Also related is the literature that seeks to describe behavior in financial markets and agricultural decision-making, by distinguishing whether agents maximize expected returns, or minimize the probability of economic disaster. This is discussed in the form of safety-first decision rules in the literature on portfolio management (Roy, 1952), and later in agricultural settings (Roumasset et al., 1976). More recently, Bryan (2017) shows that ambiguity-aversion (evaluating choices based on their minimum returns) is an important barrier to the adoption of risky technologies. However, these papers do not consider learning. Epstein and Schneider (2007) model learning under ambiguity with calibrated applications to stock-market participation. This paper differs from Epstein and Schneider (2007) in that it estimates a structural, already existing, model of robust learning from data, and contrasts the goodness-of-fit of this model with standard normal learning models.

Robust learning of the type that I consider in this paper arises from the robust control and filtering literature which originated in statistics (Whittle, 1981) and engineering (Banavar and Speyer, 1994; Shaked and Theodor, 1992), and was established in economics in the macroeconomic literature (Hansen and Sargent (2008) contains a detailed treatment of these topics). This kind of uncertainty regarding model misspecification (where the agent is uncertain that he has specified the relationship between the noisy signal and underlying state correctly) is also closely linked to the literature on ambiguity aversion, where agents operating in environments where they do not know the underlying environment evaluate returns to different actions based on each action's minimum returns (Gilboa (2009) contains a detailed decision theoretic treatment of choice under uncertainty).

2. Context

The Indian monsoon is part of a larger Asian-Pacific monsoon, which is vital to the agricultural sector and economies of several countries. In India, the four months of the monsoon (June-September) account for about 80% of annual rainfall. Figure 1 illustrates daily cumulative rainfall over the course of the year for the villages in the dataset I use (which detailed in section 5), with the start and end of the monsoon marked out - the significant role of the four monsoon months in annual cumulative rainfall is clear. Furthermore, since only about a third of arable land is irrigated in India (World Bank, 2012), monsoon risk is a crucial component of weather risk.

In particular, the timing of the onset of the monsoon (the first phase of the monsoon accompanied by an increase in rainfall relative to earlier months) is an important aspect of agricultural profits (Rosenzweig and Binswanger, 1993), since it provides the soil moisture necessary for the early stages of plant growth. There is considerable evidence that monsoon rainfall, and onset timing and intensity in particular, are affected by global climate phenomenon such as the El Niño (Kug, Jin, and An, 2009). Interestingly, these relationships have been increasingly variable in recent decades (Ummenhofer, Gupta, Li, Taschetto, and England, 2011), and there is some evidence that the progression of the monsoon within India might be slowing (Goswami, Kulkarni, Mujumdar, and Chattopadhyay, 2010). Given recent decadal changes in the monsoon, and increased awareness about climate change, it is plausible that farmers are uncertain about the exact, possibly time-varying, stochastic processes governing the timing and intensity of the monsoon.

Furthermore, decisions regarding the timing of planting are important for agricultural profits. Farmers choose an optimal planting time taking into account soil moisture (Giné, Townsend, and Vickery, 2007), the pest environment, and rainfall signals (along with possibly a variety of other signals, a case I discuss in section 6.5.1). Replanting costs are usually high (Giné, Townsend, and Vickery, 2007), and include the adverse impact of a shorter growing season. Studies in the agronomic literature that utilize plant growth models to estimate optimal sowing windows find rainfall variation at sowing to be a crucial component of yields (Rao, Gadgil, Rao, and Savithri, 2000). Thus, in addition to being an important economic decision, planting time is a measure of farmers' beliefs regarding the optimal sowing window each year, and panel data of planting decisions includes information on how households' beliefs about optimal planting time evolve over time.[3]

3. Model

Consider a farmer who chooses the time $\widehat{\eta}_t$ at which to plant his crops at each year t.[4] The actual optimal planting time at year t is denoted by η_t - it is unobserved by the farmer and is drawn from distribution G_t with mean μ_t. At the end of each agricultural season t, the farmer receives a noisy but informative signal y_t of the optimal planting time η_t in the village that year (thus y_t inform decisions $\widehat{\eta}_{t+1}$ onwards).

While I discuss in detail what the signal y_t actually corresponds to in the data in Section 5, it is worth briefly describing here as well to fix ideas. I assume that farmers observe the cumulative amount of fallen rainfall from the start of the monsoon and plant their crops at a point at which they feel the timing is optimal (there is sufficient soil moisture, fewer

[3]While other decisions such as investments in irrigation also include information on farmers' expectations regarding future weather, household-specific liquidity constraints might impact farmers' ability to invest in them. Thus, variation in decisions like irrigation might not as neatly map on to variation in future expectations, which is important to identify across decision rules.

[4]The estimation only considers farmers' *kharif* planting, which is the monsoon agricultural season.

pests, etc.). Hence, the planting time $\widehat{\eta}_t$ is expressed in terms of the amount of cumulative rainfall (as opposed to calendar time) at which the farmer chooses to plant. The signal y_t corresponds to the cumulative amount of rainfall after which planting led to maximal profits (across farmers) in the village (which farmers can observe prior to next year's planting season) in the year t.[5]

The farmer's expected profit depends on how far his chosen planting time is from the optimal and is given by

$$\pi_t(\widehat{\eta}_t, G_t) = a_t - b \int (\widehat{\eta}_t - \eta_t)^2 \, dG_t(\eta_t). \tag{1}$$

Here, all other decisions made by the household that influence agricultural profits are incorporated in the term $a_t \geq 0$ and $b \geq 0$ is a parameter that determines the sensitivity of profits to the chosen planting date. Since b does not vary over time, it is without loss to normalize $b = 1$ in what follows.[6]

This quadratic specification is analogous to the target-input model commonly employed to model agricultural profits (Conley and Udry, 2001; Foster and Rosenzweig, 1995). While relatively flexible, this functional form has the following consequences. (i) The choice of planting date η_t and the other decisions a_t (such as investments in capital, labor, fertilizer etc.) are assumed to impact profits in an additively separable way. The optimal planting time is a composite of various agronomic factors (such as the level of soil moisture to facilitate seed germination, sunlight, humidity, and the pest environment) and hence, it is realistic to assume that it does not depend on investments. (ii) The chosen planting time in period t does not affect profits in any subsequent time period. This is an appropriate assumption for this setting since the farmers in my data plant seasonal crops like cotton and rice (and not tree crops), which have to be replanted every year. (iii) Planting too early has the same (negative) marginal consequences as planting too late.[7]

I now describe the stochastic process governing the optimal planting η_t. In words, at each time t, η_t is normally distributed with mean μ_t which in turn follows a random walk. I refer to the mean of the optimal planting time μ_t as the *state*.

Formally, at time $t = 1$, the farmer's prior belief is that the mean optimal planting time μ_1 is normally distributed with mean μ_0 and variance ω^2. The state evolves according to a random walk

$$\mu_t = \mu_{t-1} + v_t, \tag{2}$$

[5] I explore robustness to alternative definitions of y_t in the data in Section 6.5.2.

[6] The theory can readily accommodate a time dependent parameter b_t. However, since the farmers' profit functions are unobserved, such a time dependent b_t cannot be identified or estimated from the data.

[7] This assumption is standard in target-input models (Conley and Udry, 2001; Foster and Rosenzweig, 1995). Since the true optimal planting time is unobserved (though farmers observe noisy signals of it every period), the assumption is not directly testable. However, if the loss of planting too early vs too late is asymmetric - for instance, $a - b_1 \mathbb{1}(\widehat{\eta}_t \geq \eta_t)(\widehat{\eta}_t - \eta_t)^2 - b_2 \mathbb{1}(\widehat{\eta}_t < \eta_t)(\widehat{\eta}_t - \eta_t)^2$, where $b_1 \neq b_2$, Granger (1969) shows that the optimal $\widehat{\eta}_t$ is $\mathbb{E}(\eta_t)$ plus a constant term that depends b_1/b_2. In neither case is the relative weighting of the past history of rainfall signals and the prior affected, which is what is crucial to the identification across learning rules.

where v_t is drawn independently across time from a normal distribution with mean 0 and variance ϕ^2. Note that the special case $\phi^2 = 0$ corresponds to a time-invariant state. The optimal planting time is normally distributed with mean μ_t or

$$\eta_t = \mu_t + \delta_t, \tag{3}$$

where δ_t is Normally distributed with mean 0.

The farmer learns about the optimal time to plant by observing *signals*

$$y_t = \eta_t + \varepsilon_t = \mu_t + \delta_t + \varepsilon_t \tag{4}$$

where the noise ε_t is independently drawn in each period and is normally distributed with mean 0. We denote the total variance of the noise terms $\delta_t + \varepsilon_t$ by σ^2. The farmer uses the signals $\{y_1, \ldots, y_{t-1}\}$ to learn the mean μ_t which in turn determines the optimal planting time η_t.

F_t denotes the distribution obtained from Bayesian updating (which will be a normal distribution the mean and variance of which depend on the signals). A non-robust farmer will maximize profits π_t with respect to this distribution or, in my notation, considers $G_t = F_t$ and solves

$$\max_{\widehat{\eta}_t} \left\{ \pi_t \left(\widehat{\eta}_t, F_t \right) \right\}.$$

It is immediate to see that the solution to the above problem is to choose a planting date equal to the mean of F_t.

By contrast, a robust farmer accounts for the fact that the true distribution G_t may differ from that obtained by Bayesian updating F_t (or that his underlying model of the world may be misspecified). Specifically, he believes that an adversarial nature picks a distribution that *minimizes* his cumulative expected profits and he chooses a planting date to maximize the *worst-case* expected profits. Formally, for a given choice of planting date $\widehat{\eta}_t$, the farmer evaluates expected profits according to a distribution $\hat{G}_t(\widehat{\eta}_t)$ that solves

$$\hat{G}_t(\widehat{\eta}_t) = \underset{G_t}{\operatorname{argmin}} \left\{ \underbrace{\pi_t(\widehat{\eta}_t, G_t) + \sum_{t'=1}^{T-1} \pi_t(\widehat{\eta}_{t'}, G_t)}_{\text{Cumulative Profits}} + \underbrace{\frac{1}{\theta} KL(G_t, F_t)}_{\text{Nature's Penalty}} \right\}, \tag{5}$$

where $\theta \geq 0$ and $KL(G_t, F_t)$ is the Kullback-Liebler divergence between the distributions G_t and F_t (that I describe in further detail below). This latter term disciplines the problem: $\frac{1}{\theta} KL(G_t, F_t)$ is interpreted as the penalty imposed on nature when she picks a distribution G_t that is different from F_t. Formally, the function

$$KL(G_t, F_t) = \int_{-\infty}^{\infty} \log \left(\frac{g_t(\eta_t)}{f_t(\eta_t)} \right) g_t(\eta_t) d\eta_t,$$

is a measure of the distance between the distributions G_t and F_t (with corresponding densities g_t and f_t).$KL(G_t, F_t) \geq 0$ for all distributions G_t, and $KL(G_t, F_t) = 0$ if and only if $g_t = f_t$ almost everywhere. Intuitively, it is a measure of the information lost when G_t is used to approximate F_t. This term disincentivizes nature from picking distributions that are arbitrarily far from the distribution F_t (thereby preventing nature from always driving the worst-case expected profits to 0).[8]

Here, θ captures the *degree of robustness*: larger values of θ impose a greater penalty on nature implying that the worst-case distribution G_t will be "closer" to F_t. Observe that as $\theta \to 0$, the penalty term $\frac{1}{\theta}KL(G_t, F_t)$ explodes implying that nature is compelled to choose $G_t = F_t$ or, equivalently, that the standard normal Bayesian learning framework is obtained as a special case (where the farmer's degree of robustness is $\theta = 0$).

A robust farmer chooses an optimal planting date $\widehat{\eta}_t$ (that maximizes worst-case profits) by solving

$$\max_{\widehat{\eta}_t} \left\{ \pi_t \left(\widehat{\eta}_t, \hat{G}_t(\widehat{\eta}_t) \right) \right\}. \tag{6}$$

Expressing both (5) and (6) together, the farmer solves the following max-min problem

$$\max_{\widehat{\eta}_t} \min_{G_t} \left\{ a_t - \int (\widehat{\eta}_t - \eta_t)^2 \, dG_t(\eta_t) + \sum_{t'=1}^{T-1} \left[a_{t'} - \int (\widehat{\eta}_{t'} - \eta_t)^2 \, dG_t(\eta_t) \right] + \frac{1}{\theta} KL(G_t, F_t) \right\}. \tag{7}$$

The choice of modeling the penalty term on nature by using the Kullback-Liebler divergence (as opposed to using a different distance function) is a convenient functional form assumption.[9] As I will describe below, this assumption combined with the assumption of normality (of the distributions of the prior, the signals and the state evolution), implies that an elegant feature of this model is that the above seemingly complex optimization problem yields a simple and intuitive solution.

I now describe in turn, the optimal planting date as a function of the observed signals for our three main cases of interest. This will yield the relevant structural equations that I estimate in the empirical Section 4.

3.1. *The (Non-Robust) Normal Learning with a Time-Invariant State*

As mentioned earlier, this is the workhorse learning model used in the empirical literature. Recall that this corresponds to the special case where $\phi^2 = 0$ i.e. that the mean optimal planting time μ_t does not change over time. Here, the period-t optimal planting time (which is the mean of the posterior distribution F_t) is given by the following linear recursive equation

$$\widehat{\eta}_{t+1} = \left(1 - K_t^N\right) \widehat{\eta}_t + K_t^N y_t,$$

where

$$K_t^N = \frac{\Sigma_t^N}{\Sigma_t^N + \sigma^2}, \qquad \Sigma_t^N = \frac{\Sigma_{t-1}^N \sigma^2}{\Sigma_{t-1}^N + \sigma^2}, \qquad \Sigma_0^N = \omega^2. \tag{8}$$

[8]An equivalent way rewrite (5) would be to remove the penalty term $\frac{1}{\theta}KL(G_t, F_t)$ and instead constrain nature to pick a distribution G_t from a *set* of distributions, each element of which cannot have more than some given maximal Kullback-Liebler divergence from F_t. The maximal distance (or equivalently the "size" of the set) would then be parametrized by θ with higher values representing a greater degree of robustness (as nature can pick from a larger set of distributions).

[9]For similar reasons, the rational inattention literature assumes that the cost of information acquisition is the reduction in entropy.

As discussed in Section 1 , an important qualitative prediction is that all signals are weighted equally (after expanding the recursion, the coefficients on $\{y_1, \ldots, y_t\}$ are the same) since the state is time-invariant and all signals are equally informative.[10]

3.2. *The Robust Learning Model*

My main model of interest is one where the underlying state is time-invariant ($\phi^2 = 0$) but where the farmer exhibits a concern for robustness. As mentioned earlier, this model might be thought to be particularly relevant for a farmer who makes weather dependent decisions in an environment where the distribution of weather is possibly changing. A strength of this framework is that the seemingly complex optimization problem (7) yields a simple and intuitive solution. Once again, the period-t optimal planting time is given by a linear recursive equation

$$\widehat{\eta}_{t+1} = \left(1 - K_t^R\right) \widehat{\eta}_t + K_t^R y_t,$$

where

$$K_t^R = \frac{\Sigma_t^R}{\Sigma_t^R + \sigma^2(1 - \theta\Sigma_t^R)}, \qquad \Sigma_t^R = \frac{\Sigma_{t-1}^R}{\Sigma_{t-1}^R + \sigma^2(1 - \theta\Sigma_{t-1}^R)}, \qquad \Sigma_0^R = \omega^2. \tag{9}$$

Contrast (9) with (8) and observe that the optimal planting time depends on the degree of robustness θ (where (8) is the special case corresponding to $\theta = 0$). Qualitatively, as with the case of a time-varying state, learning involves placing a greater weight on more recent signals but for a different reason. Intuitively, the reason for this "recency bias" is that, because the farmer uses old information in more forecasts (e.g. the first signal enters the farmer's forecast for all periods starting from the second period onwards, the second signal enters the their forecast for all periods starting from the third period, and so on), nature can maximize the farmer's estimation error by making earlier signals more noisy. Thus, in updating his beliefs, the farmer best responds by treating the earlier signals as less informative.

To make the paper self-contained (and to provide an easy reference for those readers not familiar with this model), I provide a derivation in Appendix B that shows how to reduce the min-max problem (7) to an equivalent and simpler min-max problem with a quadratic objective in which nature picks a single variable as opposed to a function G_t. Once restated this way, it is then straightforward to obtain the equation (9) for the optimal planting time (alternatively, see Simon (2006) for a derivation). Additionally, in Appendix B, I discuss an alternative (and, in my opinion, less compelling) formulation (due to Whittle (1981)) which yields the same behavior but is purely Bayesian.[11]

[10]This is not immediately apparent from the recursive formulation but can be verified by substituting $\widehat{\eta}_t = \left(1 - K_{t-1}^N\right) \widehat{\eta}_{t-1} + K_{t-1}^N y_{t-1}$ and so on and by observing that $(1 - K_t^N)K_{t-1}^N = K_t^N$.

[11]Essentially, Whittle (1981) demonstrates that it is possible to construct a particular utility function for the farmer which yields the behavior described by (9). This utility function is somewhat hard to interpret (in that there is not compelling interpretation for this particular function) which is why I find it less compelling as a model.

3.3. *The (Non-Robust) Normal Learning Model with a Time-Varying State*

When the mean optimal planting time μ_t follows a random walk (given by (2)) with $\phi^2 > 0$, the period-t optimal planting time is once again given by a linear recursive equation

$$\widehat{\eta}_{t+1} = \left(1 - K_t^{RW}\right)\widehat{\eta}_t + K_t^{RW} y_t. \tag{10}$$

where

$$K_t^{RW} = \frac{\Sigma_t^{RW}}{\Sigma_t^{RW} + \sigma^2}, \qquad \Sigma_t^{RW} = \frac{\Sigma_{t-1}^{RW}\sigma^2}{\Sigma_{t-1}^{RW} + \sigma^2} + \phi^2, \qquad \Sigma_0^{RW} = \omega^2. \tag{11}$$

Contrast (11) with (8) and observe that $K_t^{RW} \neq K_t^N$. It is no longer the case that all signals are equally weighted; indeed, upon expanding the recursion it can be verified that that the coefficients on more recent signals are greater (that is, the coefficient on y_t will be greater than y_{t-1} and so on). This is intuitive: since the state is time-varying, more recent signals contain less noisy information about the underlying state.[12]

3.4. *Summary of the Theoretical Framework*

To summarize, in all three models, updating is linear (a feature that proves very useful for estimation and model selection). The optimal choice $\widehat{\eta}_{t+1}$ is obtained from the optimal choice in the previous period $\widehat{\eta}_t$ and the signal y_t via the recursive equation

$$\widehat{\eta}_{t+1} = (1 - K_t)\,\widehat{\eta}_t + K_t y_t,$$

where K_t, the *Kalman gain* at time t, is the weight given to the signal at time t. Expanding the above recursion allows for an expression of $\widehat{\eta}_{t+1}$ in terms of the entire vector of signals and the prior as

$$\widehat{\eta}_{t+1} = \widehat{\eta}_1 \prod_{j=1}^{t}(1 - K_j) + K_1 y_1 \prod_{i=2}^{t}(1 - K_i) + K_2 y_2 \prod_{i=3}^{t}(1 - K_i) + \cdots + K_t y_t. \tag{12}$$

In Figures 3-5, I demonstrate graphically the qualitative features of each of the three models by plotting the relative weights placed on past signals. Each of these figures examines a planting decision at time $t = 11$ and plots the coefficients of the signals $y_1, \ldots, y_{10}$ (from equation (12)) where the x-axis corresponds to the signal time (more recent signals are to the right).

Figure 3, which considers the NLM with a time-invariant state, shows that increasing σ^2 shifts the weight on information downward (relative to the weight on the prior), as a higher variance of the signal implies that the signals are less informative. Additionally, it shows that all the signals are weighted equally. Conversely, for the NLM with a time-varying state, Figure 4 illustrates that, as the variance of the state ϕ^2 increases, the relative importance of recent information is greater (as previous signals provide less precise information about the current state). Finally, Figure 5 demonstrates the recency bias that arises in the robust learning framework (despite the state remaining unchanged) and its increase in the degree of

[12] A robust random walk learning model would be a generalized version and nest these three models. In the robustness checks, I estimate the model and show that it does not seem to be best explaining the data.

robustness. Additionally, the difference in the shapes of the curves corresponding to Figures 5 and 4 highlight the qualitative difference in the recency bias that arises in both these cases.

The main contribution of this paper is to structurally estimate (using the equations governing the optimal planting time) the parameters corresponding to each of the three models and determining which of them best explains the behavior of the farmers in my data. For convenient reference, Table 1 summarizes the Kalman gains, K_t and the variance of the estimation error, Σ_t in each of the three cases.

4. Estimation

Recall that, in my model, a profit maximizing farmer chooses his planting date as a linear combination of his prior belief and the signals he has received. Specifically, the optimal planting time $\widehat{\eta}_{it}$ chosen by farmer i in year t depends on the past village level yearly rainfall signals $\{y_{v1}, \dots, y_{vt-1}\}$ as follows

$$
\begin{aligned}
\widehat{\eta}_{i2} =& (1-K_1)\widehat{\eta}_{i1} + K_1 y_{v1}, \\
\widehat{\eta}_{i3} =& (1-K_2)(1-K_1)\widehat{\eta}_{i1} + (1-K_2)K_1 y_{v1} + K_2 y_{v2}, \\
&\vdots \\
\widehat{\eta}_{iT} =& (1-K_{T-1})(1-K_{T-2})...(1-K_1)\widehat{\eta}_{i1} + (1-K_{T-1})(1-K_{T-2})\cdots K_1 y_{v1} \\
&+ (1-K_{T-1})(1-K_{T-2})\cdots K_2 y_{v2} + \cdots + K_{T-1} y_{vT-1}.
\end{aligned}
$$

The weight K_{t-1} that the farmer places on his most recent signal depends on the specific learning model being estimated (Table 1 contains the explicit functional forms). My estimation procedure can be described as follows. Each of my three main models of interest contain different parameters (summarized in Table 2) which I separately estimate from the above equations by using non-linear least squares regression.[13]

After estimating each of the learning rules separately, I compare the relative fit of the models using three goodness-of-fit measures: the Akaike Information Criteria (AIC), the Akaike Information Criteria corrected for smaller samples (AICc), and the Bayesian Information Criteria (BIC). All three information criteria trade-off model complexity against goodness-of-fit, and penalize additional parameters added to a model, with the BIC penalizing additional parameters relatively more. This latter step is important as both the robust NLM and the NLM with a time varying state have additional parameters (compared to the NLM with a time invariant state) which must be accounted for when determining which of the models best explains the data (and these information criteria trade-off model complexity against goodness-of-fit).

A few specific remarks on the estimation procedure are in order.

- Note that due to the recursive structure of the model, the observed planting date $\widehat{\eta}_{i1}$ in year 1 contains all the relevant unobserved information the farmer possesses prior to the start of my data. Put differently, my model allows me to use this planting date to capture the prior belief of the farmer.

[13]The parameters were estimated using the "nlinfit" tool in Matlab. Initial values were set at 0.5.

- Observe that the Kalman gains K_t (Table 2) depend on the *ratio* of the variances of the prior belief, the signal noise and the state evolution process. Hence, parameters can only identified up to scale. Therefore, I normalize the variance of the prior belief ω^2 to 1 in all the learning models and recover the rest of the parameters up to scale. Note that, following this normalization, all remaining parameters are identified.

5. Data

5.1. *Household Data*

I use household-level panel data from the Indian Crop Research Institute for the Semi-Arid Tropics (ICRISAT). The data contain detailed socio-economic information, including, season-level data on agricultural operations' timing, costs and returns. The data cover six villages over 2005-2012, and eleven additional villages over 2009-2012. The villages are from five Indian states - Andhra Pradesh, Gujarat, Karnataka, Maharashtra, and Madhya Pradesh. Thus, there is considerable cross-sectional variation in the timing of the monsoon progression, as evinced by Figure A1, which shows the location of the villages in the data.

A household's optimal planting time (corresponding to $\widehat{\eta}_t$ in the model) is chosen to be the point in the monsoon (kharif) season at which they begin planting their most important crop (the crop that has the highest contribution to profits).[14] the following points are important to re-iterate in this section (i) The planting "time" is measured to be the amount of cumulative rainfall (that has fallen since the start of the monsoon) at which planting begins. This choice reflects the fact that cumulative rainfall (instead of calendar time) determines the conditions (such as the level of soil moisture, the amount of pests etc.) that are conducive to planting (although the empirical results are qualitatively similar if I use calendar time instead). This is also consistent with the literature studying farmers' beliefs about monsoon rainfall in India- for instance, Giné, Townsend, and Vickery (2015) find that farmers are more likely to understand and respond to monsoon onset as cumulative rainfall rather than calendar time. (ii) As argued in Section 3, the chosen planting time contains all the relevant information about a household's expectation of the (unknown) optimal planting time (which thereby obviates the need to explicitly elicit farmer's beliefs). This feature (inferring beliefs from actions) is common to the literature on testing across learning models (see, for instance, Camerer and Ho (1999)).

Summary statistics for the plot on which households planted their most important crop are presented in Table 3. On average, households wait until cumulative rain after June 1 is about 120 mm, although there is significant heterogeneity across households and years. Table 1 also shows plot-level asset values, profits and area. Plots on average are about 2.14 acres, and are valued at about Rs. 167,221.70. Average profits for the kharif season for these plots are about Rs. 9,811.74.

In the learning model estimation, I use households for which the data contains 2 or more periods of information on planting dates (since at least one updating equation must exist for

[14] The results are qualitatively similar if I use the household's first planting, regardless of crop, as their expectation of the optimal planting time.

the household to be included in the estimation). If a household has gaps in years for which planting data is available, I use all the available years since the data has rainfall signals for every year and therefore, I observe the signals for the household even when I do not observe their planting date.

5.2. *Rainfall Data*

The rainfall data used in the estimation are from a precipitation data product known as CPC Morphing Technique ("CMORPH"). The data are produced by combining precipitation estimates from several satellite sources, and are available at the 3-hourly temporal resolution, and the 0.25 by 0.25 degree spatial resolution (Joyce, Janowiak, Arkin, and Xie, 2004). I use village geographic coordinates to assign the nearest CMORPH grid point to each village. The CMORPH data range from 2003-2012.

I define the signal of the optimal planting time (corresponding to y_t in the model) as the cumulative rainfall after which planting led to the maximal average profits (across farmers) in the village last year. For instance, if planting between 26th-29 June led to highest average profits last year, the cumulative rainfall that fell from the start of the monsoon (that is, from June 1) is the rainfall signal this year. I show in Section 6.1 that it predicts household planting behavior, and explore robustness to alternative definitions in Section 6.5.2.

As Table 3 shows, the average amount of cumulative rain after June 1 following which planting leads to maximal average profits (across farmers in a given village for a given year) is about 120 mm. Figure 2 presents a kernel density of planting times relative to the rainfall signal, which is described in Section 5.2. Since the CMORPH data is not available before 2003, it cannot be used to plot and distinguish the distributions of optimal planting time from previous decades. For these years, I instead use the rainfall data by the Indian Meteorological Department (IMD), which is daily station-level rainfall data at the 1° by 1° level from 1951-2007.[15]

6. Results

6.1. *Reduced Form Evidence*

I begin by showing that farmers' planting decisions are affected by the rainfall signal. To do so, I estimate the following regression:

$$pd_{ivt} = \alpha + \beta s_{vt} + \delta_i + \gamma_t + \psi_{ivt} \tag{13}$$

where pd_{ivt} is the cumulative rainfall (after June 1) following which household i in village v plants their most important crop at year t.[16] s_{vt} is cumulative planting rainfall that maximized average profits in the village at time $t-1$ (defined as the signal at time t, and varies at the village by year level). The above specification includes household fixed effects (δ_i) and year fixed effects (γ_t), but I also estimate three additional specifications - no control variables, year fixed effects, and village and year fixed effects.

[15] Since it is not available after 2007, and household data is available until 2012, I cannot use this data for the main estimation.

[16] Recall that the most important crop is the one that contributes most to profits.

Table 4 present the results of the above specifications. In all specifications, a higher level of profit-maximizing planting rainfall in a given year causes households to wait for greater cumulative rainfall before planting in the next year. The coefficient ranges from 0.22 to about 0.07 mm— so a 1mm higher profit-maximizing planting rainfall one year causes households to wait for an extra 0.07-0.22 mm of cumulative rainfall before planting next year. The estimates are statistically significant across specifications.[17]

This evidence suggests that the rainfall signal I consider is a good predictor of the households' planting decision. In the next section, I estimate the structural parameters in the farmers' belief updating equations and test which model best fits farmer behavior.

Furthermore, I show that deviation from the optimal planting time is harmful for agricultural profits. To do so, I regress the inverse hyperbolic sine of agricultural profits at the farmer-year level on the log of the squared deviation from the optimal planting rainfall in the village that year. Results are presented in Table A2. Column 1 shows this relationship without any controls, column 2 with village and year fixed effects, column 3 with village by year fixed effects, and column 4 with village by year fixed effects as well as household fixed effects. Across specifications, a greater deviation from the optimal planting time implies lower profits. The coefficient ranges from -0.135 to -0.241, which is about 2.25 to 4% lower profits relative to mean profits of a 1% deviation from the optimal time. The results are statistically significant across specifications. Thus, the timing of this agricultural decision is important for profits, and small deviations from the optimal time lead to lower profitability.

Table A3 additionally presents each reduced form version of the updating model across each year of data, which I estimate structurally as a set of equations. It regresses planting rainfall on the farmer's prior belief as well as all past signals - column 1 presents this for the third year of data, with the independent variables being the prior belief (proxied for by the planting rainfall in the first year) and the rainfall signal in the second year. Column 2 presents this for the fourth year of data, which includes the rainfall signal in the third year, and so on with each column adding in an additional year of signals. The results broadly show greater weight on more recent signals.

6.2. *Structural Learning Model Parameters*

In this section, I test which of the three models best fits household planting timing behavior. As detailed in Section 4, I use nonlinear least squares to estimate the structural parameters, and goodness of fit measures (the Akaike Information Criteria, the Akaike Information Criteria corrected for smaller samples, and the Bayesian Information Criteria) to select across models. Also as mentioned in Section 4, the variance of the prior ω^2, a parameter that appears in all three models, is normalized to 1; all the other parameters are identified up to scale. Having normalized ω^2, I separately estimate the following parameters for each model. The variance of the rainfall signal σ^2 appears in all three models. The two

[17]Since rainfall in the previous year might increase soil moisture in the current year and therefore affect planting decisions via a moisture overhang effect, I can re-estimate all four regressions controlling for one-period lagged monsoon rainfall. The coefficients on the rainfall signal are nearly identical in magnitude and statistical significance to the specification omitting lagged rainfall.

more general models each have an additional distinct parameter: the variance ϕ^2 of the state evolution in the NLM with a time-varying state and the degree of robustness θ in the robust NLM.

The structural parameter estimates are given in Table 5. In the NLM, σ^2 is estimated to be about 0.79, with a standard error of about 0.10. The robust NLM specification estimates σ^2 to be about 1.2, and θ of about 0.24. The random walk specification estimates σ^2 to be 0.83 and ϕ^2 to be 0.02 - however, I am unable to reject that ϕ^2 equals 0. Recall that $\phi^2 = 0$ corresponds to the NLM with a time-invariant state and so this indicates that farmers' actions do not appear to be consistent with the beliefs that the optimal planting time evolves as a random walk.

In contrast, we are easily able to reject the hypothesis that θ equals 0 in the robust NLM specification, which indicates that farmers are exhibiting some degree of robustness when learning about the optimal time to plant. To reiterate, this evidence is consistent with farmers exhibiting a concern about model misspecification when trying to learn about a weather dependent decision in an environment made uncertain by potential climate change.

Three goodness of fit measures - the Akaike Information Criteria (AIC), a version of the AIC that corrects for a small sample (AICc), and the Bayesian Information Criteria (BIC) are also reported in Table 5. Like the parameter estimates, the AIC measures indicate a preference for the robust NLM specification (although the evidence from the BIC does not show a preference between the normal and robust model). Thus, the parameter values and the AIC measures indicates that the robust NLM fits farmers' planting decisions best. To the best of my knowledge, this constitutes the first empirical evidence of robust learning behavior in micro-data and is one of the contributions of this paper.

6.3. *Impact of Changing Weather Distributions on the Degree of Robustness*

In this section, I further build on the evidence of robust learning (from Section 6.2) by examining learning heterogeneity based on past weather distributions. Since robustness is a way for farmers to protect themselves against uncertainty (of the true stochastic process determining the weather), it is reasonable to expect that farmers who have experienced greater changes in the weather distribution to be more robust. Put differently, a farmer who is exposed to relatively different recent weather distributions relative to the past might exhibit a greater concern for model misspecification relative to a farmer exposed to very similar weather distributions over time.

I test this hypothesis by separately estimating the learning model parameters for villages that have experienced greater and lesser changes in the distribution of the rainfall signal in previous decades. To measure changes in monsoon onset distributions, I use IMD rainfall data and divide the sample into two equal time periods of 27 years each (1951-1978 and 1979-2004) before 2005, which is the first year for which planting data is available. Recall that my measure of the onset signal in a given year was the level of cumulative rainfall planting in which maximized average profits in the village the previous year. This is not available for years in which household-data is not available, and so I set the village-level rainfall signal to be the mean profit-maximizing rainfall level in that village, and estimate the empirical distribution of the timing (at the pentad, or 5-day, level) of this rainfall signal for each village for each of these two time periods. Then, for each village, I calculate the difference between the earlier (1951-1978) and recent (1979-2004) signal distributions by computing the Kullback-Leibler divergence (henceforth, referred to as KLD) between them. Recall from Section 3 that the KLD provides a measure of the distance between any two distributions where greater values correspond to distributions that are more "different." Specifically, if the value of the KLD is 0, it means that these two distributions are identical (no information is lost by using the earlier distribution to approximate the recent one). Conversely, greater values of the KLD imply that the recent distribution is more unlike the earlier one.

I divide the sample into villages with above and below median value of the KLD; the interpretation is that the former (latter) are villages which have experienced greater (lesser) changes in the recent rainfall signal distributions compared to the past. Figure 6 presents two examples from each of the samples. The village on the left has a high KLD and is in the above median sample, whereas the village on the right has a lower KLD and is in the below median sample. A superficial inspection of Figure 6 shows that the distributions pre and post 1978 in the low KLD village appear quite similar whereas the post 1978 distribution appears different from the pre-1978 distribution for the high KLD village. The figure also suggests that the distribution of the high KLD village has greater variance and so it is worth emphasizing that the KLD measure is not driven by variance; indeed, two identical distributions with the same high variance would have a KLD of 0.

Results of the learning model estimation conducted separately on villages with high (Panel A) and low KLD (Panel B) are given in Table 6. For the high KLD villages, the degree of robustness θ is large (0.42), and statistically significant. By contrast, for the low KLD villages, θ is not only much smaller (0.05) but statistically not different from 0. Put together, these intuitive results (farmers in villages with historically more volatile signal distributions are more concerned about whether they know the true underlying stochastic process) provide still further evidence of robust learning.

The main implication of the robust learning model, as I have discussed above, is that the farmer will minimize worst-case (maximum) forecast errors. To test if this is true in the raw data, I construct a forecast error for each farmer each year, which is the square of the deviation of the farmer's planting rainfall relative to the optimal planting rainfall in their village that year. With these forecasts, I construct a farmer level measure of the greatest forecast error (the maximum of their forecast errors across all periods). I plot these raw data by whether a village experienced a change in the distribution (high KLD) vs. it did not (low KLD) in Figure 7. The density of maximum forecast errors is shifted leftward in villages with a high KLD, indicating that farmers in these villages have lower worst-case (maximum) forecast errors.

Table 9 presents regression results that show this relationship, controlling for mean squared forecast error, and also present results for minimum forecast errors (the latter as a placebo check). Columns 1 presents the results from a cross-sectional regression of the maximum forecast error by a farmer on a dummy variable that takes the value 1 if the village is a high KLD village, and 0 otherwise. Column 2 controls for mean forecast error. There is a highly statistically significant and large (relative to the mean) effect showing that farmers in high KLD villages have lower maximum forecast errors, which is robust to controlling for the farmer's overall mean forecast error. Columns 3 and 4 show the results from the same specification as in Columns 1 and 2 respectively, but with minimum squared forecast error as the dependent variable. In contrast, we find no relationship between the minimum forecast error and whether the village is a high KLD village, and a positive relationship once mean forecast error is controlled for. These results are important because they show that the raw data support an important implication of the the robust learning model (that farmers minimize worst-case forecast errors) which is consistent with the results from the structural estimation - farmers in villages that experienced changes in the rainfall distribution are shown to be using robust updating rules in the structural estimation, and their worst-case forecast errors are lower in the raw data.

6.4. *Learning Heterogeneity by Irrigation Access*

The previous section highlighted the heterogeneity in the degree of robustness depending on the volatility of past weather distributions. In this section, I provide further evidence in support of the model by testing how the access to irrigation (which has the potential to imperfectly insure the farmers against extreme monsoon realizations) changes planting behavior. I consider households that had irrigation in at least some years and did not have irrigation in other years.[18] This ensures that other time-invariant household characteristics are not affecting the results in this sub-sample.

Recall that the robust farmers solve the optimization prtance of the plot to the nearest source of irrigation, and distance of the plot from the farmer's house. This estimates the predicted probability that a ploblem (7) which can be equivalently written (by dividing through by the constant b) as

$$\max_{\widehat{\eta}_t} \min_{G_t} \left\{ \frac{a_t}{b} - \int (\widehat{\eta}_t - \eta_t)^2 \, dG_t(\eta_t) + \sum_{t'=1}^{T-1} \left[\frac{a_{t'}}{b} - \int (\widehat{\eta}_{t'} - \eta_t)^2 \, dG_t(\eta_t) \right] + \frac{1}{b\theta} KL(G_t, F_t) \right\}.$$

Recall also that I had normalized the parameter b (which captures the sensitivity of profits to the choice of planting date) to 1 as it is clearly seen from the above expression that θ cannot be separately identified from b.

Now note that the access to irrigation makes the farmer less dependent on rainfall. In the above expression, this would correspond to a lower sensitivity of profit to the choice of planting date or a reduced b. Hence, if farmers were indeed solving the above problem, we should expect that the estimated value of θ after the normalization $b = 1$ (which is actually the product $b\theta$ without the normalization) should be lower in years where the farmer has access to irrigation (as we expect the b is lower). Put differently, while the farmer's learning behavior from past information (captured by his *actual* θ) should not depend on his access to irrigation, the *estimated* degree of robustness should (as it also contains the parameter b).

Table 7 presents the results of the estimation separately for irrigated and unirrigated years. As we should expect from the above argument, the degree of robustness is positive

[18] Thus, households that always or never had irrigation are omitted.

and statistically significant for both sets of years. Moreover, it is much higher in unirrigated years ($\theta = 0.37$) than in irrigated years ($\theta = 0.17$). The goodness-of-fit measures across the three models are not very different, although the robust NLM is weakly preferred in unirrigated years and the NLM with time-invariant state is weakly preferred in irrigated years. It is worth pointing out that, since the estimation requires at least two observations in which the household either had irrigation or did not to be included in the sample (to form at least one updating equation), the samples are much smaller than the previous estimations.[19]

6.5. *Robustness Checks*

6.5.1. *The Econometrician only Observes a Subset of the Signals used by the Farmer*

It is possible that farmers use multiple weather signals (which are not observed by the econometrician) in making planting decisions in addition to the rainfall signal I observe. Note that this is a potential concern in all learning model estimation studies as the econometrician may only observe a subset of the signals driving decision making. That said, it is important to ensure that, even when this is the case, the estimates I obtain (of the weight placed on information from different periods) are unbiased.

In Appendix A, I show that when the farmer uses two signals of which the econometrician observes only one, the covariance between the farmer's estimate of the underlying state and the econometrician's estimate of the farmer's estimate of the underlying state is zero, that is, the time path of the decision weights is unbiased. This is true even when the error terms on the two signals are correlated. Thus, the fact that farmers might use multiple signals to infer optimal planting times, only one of which I observe, does not impact the identification. Of course, the observed signal should be a predictor of planting decisions, which has previously been shown in section 6.1.

6.5.2. *Crop Specific Rainfall Signal*

While the rainfall signal used in the main estimation predicts household behavior on average, since farmers are growing different crops in the data, I investigate the robustness of my estimates to the alternate choice of crop-specific rainfall signals. That is, instead of using the cumulative rainfall corresponding to maximized average profits (across crops), one can define the signal to be the cumulative rainfall (up till the 5-day window) after which planting maximized average profits last year for the specific crop that the household planted this year. There is a large amount of heterogeneity in the crops that form the most important crop in the ICRISAT sample— cotton is the most common (23% of plots), followed by soybean (20%), pigeonpea (17%), and maize (10%). The rest of the sample comprises crops like

[19] It is also possible to conduct a similar exercise in which learning rules are estimated separately for households with high and low wealth. Here, one would expect richer households to care less about worst-case profits and this is corroborated by results which show that the behavior of farmers with below (above) median land values is best explained by the robust NLM (standard NLM with time-invariant state). I consider these results slightly less interesting since it is difficult to separate the many possible mechanisms (access to credit, better forecasting ability, and more insurance mechanisms to name a few), and so do not include them in the main paper. They are reported in Appendix **??**.

sorghum, greengram, blackgram and millets. For this reason, and since we need a household to cultivate a crop that contributes most to their profits at least twice to be included in the estimation (since at least two observations are required to form an updating equation), the number of observations in this estimation is significantly lower than the entire sample.

Table 8 presents the results. While the variance of the signal (σ^2) is greater than when the signal is not crop-specific, the value of θ is very similar (0.22), and is still statistically significant. As in the general estimation, I fail to reject that the variance of the state evolution process (ϕ^2) is statistically significantly different from zero. Furthermore, the AIC measures indicate a weak preference for the robust NLM as in the main estimation. The overall results are thus similar as in the general estimation shown in Table 5.

6.5.3. *Alternative Models*

In addition to the three decisions tested in the previous section, a natural fourth possibility is that of a the robust random walk model. Here, the farmer solves (7) as before but he believes that the underlying state follows a random walk (that is, $\phi^2 > 0$). Once again, the farmer's optimal choice of plating time follows the same linear equation (12) with the corresponding parameters given by

$$K_t = \frac{\Sigma_t}{\Sigma_t + \sigma^2(1 - \theta\Sigma_t)} \text{ and } \Sigma_t = \frac{\Sigma_{t-1}\sigma^2}{\Sigma_{t-1} + \sigma^2(1 - \theta\Sigma_{t-1})} + \phi^2.$$

(Note the intuitive similarity to equations (11) and (9).) The fact that I fail to reject that the state is time-invariant ($\phi^2 = 0$) for the NLM already provides suggestive evidence that farmers do not perceive the state to be changing. An unconstrained estimation for the robust random walk model conducted on the main sample results in a negative value for the variance of the state evolution (ϕ^2), which is clearly outside the permissible parameter value space. Non-linear optimization restricting the parameters to be nonnegative yields an estimate of 0 for ϕ^2. Thus, the additional degree of freedom in the robust random walk model does not better explain the observed choices of farmers' planting decisions.

Secondly, there is a possibility that the evolution is a more general AR(1) process $\mu_t = \zeta\mu_{t-1} + v_t$, where $\zeta = 1$ corresponds to the time-varying NLM when the parameter of interest, the mean optimal planting time, follows a random walk. In fact, an even more general process process, given by $\mu_t = \zeta_t\mu_{t-1} + v_t$, is also possible. This is especially possible if farmers perceive the optimal planting time to follow a cyclical process for instance. I re-estimate this general NLM model allowing ζ_t to be estimable, time-varying parameters instead of fixing it to 1. The average estimated value across is $\zeta = 0.87$, and the estimated value of σ is about 0.8.

Finally, and needless to say, the set of models I estimate and compare are not exhaustive which, of course, is a limitation of estimating any structural model (as the existence of a better alternate model can never be disproved). The purpose of this study is to develop an empirical framework for robust learning, a model which has numerous advantages: it is theoretically well founded, yields natural behavioral predictions, is simple to estimate and, importantly, nests the canonical NLM. Furthermore, as indicated by the empirical results, this model is a good representation of the kind of Knightian uncertainty generated by climate change, while also being consistent with the raw data, and thus studies of adaptation should consider this model in estimating adaptive behavior.

6.6. *Back of the Envelope Calculations About the Performance of Learning Rules*

I end this section with some suggestive evidence that shows that robust learning can, at times, also lead to the most accurate forecasts, depending on whether and how much the environment the agents operate in is changing over time. While I have shown that the farmers' choice of planting times are best explained by a robust NLM, this does not necessarily imply (as robust farmers maximize worst case profits) that they were planting at the profit maximizing level of rainfall evaluated ex post at the village-level. For villages that experienced bigger changes in the historical distribution of the rainfall signal (which are also villages that have more robust farmers), I evaluated the posterior prediction of optimal planting rainfall for each year starting 2007 using the 2005 profit-maximizing planting rainfall as the prior and the 2006 profit-maximizing planting rainfall as the first signal. I then computed the mean squared error of the predictions relative to the actual levels of rainfall after which planting led to maximal profits. The robust NLM has the least mean squared error (13753.38) relative to either the NLM with time invariant (14888.33) or time varying state (14512.67). Thus, in uncertain environments, making decisions robust to model misspecification (evaluating actions from a set of possible distributions) *can* perform better even on average relative to relying solely on one model (the baseline rainfall distribution), since the latter does not incorporate the uncertainty in the environment. The relative performance of these rules across settings of course depends on the parameters of the rule itself as well as the environment, and its rate of change.

7. Conclusion

Adaptation to climate change presents one of the major challenges of the 21st century. The profits and output of farmers, especially those in the developing world, depend critically on the weather and the agricultural decisions they make in response. In particular, their capability to adapt and the nature of their adaption depends on their ability to learn about and predict the weather conditions they will face, and climate change will add additional, unknowable (in the short-term), uncertainty to this problem. In light of this, it is important to understand their decision making process as this can help identify the most vulnerable farmers, inform adaptation policies (by prioritizing specific information and resources to enhance learning) and pinpoint the most effective adaptation measures.

This paper develops an empirical framework (to test the model and estimate the underlying parameters driving learning) and finds support for the model in long-term panel data from India on monsoon realizations and household-level planting times. A notable strength of the robust learning model is that it generalizes the workhorse normal learning setting to incorporate the kind of Knightian uncertainty generated by climate change, by weakening the strong required informational assumptions while simultaneously maintaining the tractability. I show that the planting decisions of the farmers in my sample (made in response to weather signals) fits a robust learning framework better than the standard normal learning model with either a time-invariant or time-varying state (to the best of my knowledge, this paper is also the first to estimate the latter model in micro data and test it across other canonical models). Examining the heterogeneity in learning lends further support to the model, and shows that the results from the structural estimation match patterns in the raw data. The degree of robustness is more pronounced in villages which, in the last few decades, experienced a bigger change in the distribution of the rainfall after which planting led to maximal profits. In the raw data, farmers in these villages also have lower maximum forecast errors (which is predicted by the model), a relationship that does not hold for minimum forecast errors.

As discussed in Section 6.6, in this case, robust learning allows for better adaptation to uncertainty. Testing the relative benefits of other adaptive mechanisms to climate change such as irrigation and migration opportunities, as well as how household beliefs about the climate impact these decisions, remain interesting questions for future work.

References

Banavar, R. N., and J. L. Speyer (1994): "Risk-sensitive estimation and a differential game," *Automatic Control, IEEE Transactions on*, 39(9), 1914–1918.

Barham, B. L., J.-P. Chavas, D. Fitz, V. Ríos-Salas, and L. Schechter (2014): "Risk, learning, and technology adoption," *Agricultural Economics*.

Bryan, G. (2017): "Ambiguity and insurance," *Unpublished manuscript*.

Burke, M., and K. Emerick (2013): "Adaptation to climate change: Evidence from US agriculture," *University of California, Berkeley. http://www. ocf. berkeley. edu/˜ kemerick/burke_emerick_2013. pdf*.

Camerer, C., and T. H. Ho (1999): "Experience-weighted Attraction Learning in Normal Form Games," *Econometrica*, 67(4), 827–874.

Cheung, Y.-W., and D. Friedman (1997): "Individual learning in normal form games: Some laboratory results," *Games and Economic Behavior*, 19(1), 46–76.

Conley, T., and C. Udry (2001): "Social learning through networks: The adoption of new agricultural technologies in Ghana," *American Journal of Agricultural Economics*, pp. 668–673.

Conley, T. G., and C. R. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, pp. 35–69.

Deressa, T. T., R. M. Hassan, C. Ringler, T. Alemu, and M. Yesuf (2009): "Determinants of farmers choice of adaptation methods to climate change in the Nile Basin of Ethiopia," *Global Environmental Change*, 19(2), 248–255.

Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *The Quarterly Journal of Economics*, pp. 643–669.

Epstein, L. G., and M. Schneider (2007): "Learning under ambiguity," *The Review of Economic Studies*, 74(4), 1275–1303.

Ferro, C. A., A. Hannachi, and D. B. Stephenson (2005): "Simple nonparametric techniques for exploring changing probability distributions of weather," *Journal of Climate*, 18(21), 4344–4354.

Foster, A. D., and M. R. Rosenzweig (1995): "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of political Economy*, pp. 1176–1209.

Gibbons, R., L. F. Katz, T. Lemieux, and D. Parent (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 23(4), 681–724.

Gilboa, I. (2009): *Theory of decision under uncertainty*, vol. 1. Cambridge university press Cambridge.

Giné, X., R. Townsend, and J. Vickery (2015): "Forecasting when it matters: evidence from semi-arid India," *Manuscript, FRB of NY*.

Giné, X., R. M. Townsend, and J. Vickery (2007): "Rational expectations? Evidence from planting decisions in semi-arid India," *Manuscript. World Bank, Washington, DC*.

Goswami, B., J. Kulkarni, V. Mujumdar, and R. Chattopadhyay (2010): "On factors responsible for recent secular trend in the onset phase of monsoon intraseasonal oscillations," *International Journal of Climatology*, 30(14), 2240–2246.

Granger, C. W. (1969): "Prediction with a generalized cost of error function," *OR*, pp. 199–207.

Hansen, L. P., and T. J. Sargent (2008): *Robustness*. Princeton university press.

Hyslop, D. R., and G. W. Imbens (2001): "Bias from classical and other forms of measurement error," *Journal of Business & Economic Statistics*, 19(4), 475–481.

IPCC (2013): *Working Group I contribution to the IPCC 5th Assessment Report "Climate Change 2013: The Physical Science Basis". Final Draft Underlying Scientific-Technical Assessment*. Cambridge University Press.

Jacoby, H. G., and E. Skoufias (1998): "Testing theories of consumption behavior using information on aggregate shocks: Income seasonality and rainfall in rural India," *American Journal of Agricultural Economics*, 80(1), 1–14.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie (2004): "CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *Journal of Hydrometeorology*, 5(3), 487–503.

Kelly, D. L., C. D. Kolstad, and G. T. Mitchell (2005): "Adjustment costs from environmental change," *Journal of Environmental Economics and Management*, 50(3), 468–495.

Kug, J.-S., F.-F. Jin, and S.-I. An (2009): "Two types of El Niño events: cold tongue El Niño and warm pool El Niño," *Journal of Climate*, 22(6), 1499–1515.

Kurukulasuriya, P., N. Kala, and R. Mendelsohn (2011): "Adaptation and climate change impacts: a structural Ricardian model of irrigation and farm income in Africa," *Climate Change Economics*, 2(02), 149–174.

Kurukulasuriya, P., and R. Mendelsohn (2008): "Crop switching as a strategy for adapting to climate change," *African Journal of Agricultural and Resource Economics*, 2(1), 105–126.

Lemoine, D. (2018): "Estimating the Consequences of Climate Change from Variation in Weather," *NBER Working Paper*, (w25008).

Lybbert, T. J., C. B. Barrett, J. G. McPeak, and W. K. Luseno (2007): "Bayesian herders: Updating of rainfall beliefs in response to external forecasts," *World Development*, 35(3), 480–497.

Miller, B. M. (2013): "Does Validity Fall from the Sky? Observant Farmers, Exogenous Rainfall, and Climate Change," *Manuscript, UCSD*.

Rao, N. K., S. Gadgil, S. P. Rao, and K. Savithri (2000): "Tailoring strategies to rainfall variability-The choice of the sowing window," *Current Science*, 78(10), 1216–1230.

Rosenzweig, M., and C. R. Udry (2013): "Forecasting profitability," Discussion paper, National Bureau of Economic Research.

Rosenzweig, M. R., and H. P. Binswanger (1993): "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *Economic Journal*, 103(416), 56–78.

Roumasset, J. A., et al. (1976): *Rice and risk. Decision making among low-income farmers.* North Holland Publ. Comp.

Roy, A. D. (1952): "Safety first and the holding of assets," *Econometrica: Journal of the Econometric Society*, pp. 431–449.

Shaked, U., and Y. Theodor (1992): "H-optimal estimation: A tutorial," in *Decision and Control, 1992., Proceedings of the 31st IEEE Conference on*, pp. 2278–2286. IEEE.

Simon, D. (2006): *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.

Taraz, V. (2017): "Adaptation to climate change: Historical evidence from the Indian monsoon," *Environment and Development Economics*, 22(5), 517–545.

Ummenhofer, C. C., A. S. Gupta, Y. Li, A. S. Taschetto, and M. H. England (2011): "Multi-decadal modulation of the El Nino–Indian monsoon relationship by Indian Ocean variability," *Environmental Research Letters*, 6(3), 034006.

Whittle, P. (1981): "Risk-sensitive linear/quadratic/Gaussian control," *Advances in Applied Probability*, pp. 764–777.

World Bank (2012): *World development indicators*. World Bank.

Figure 1. Cumulative Precipitation During the Year

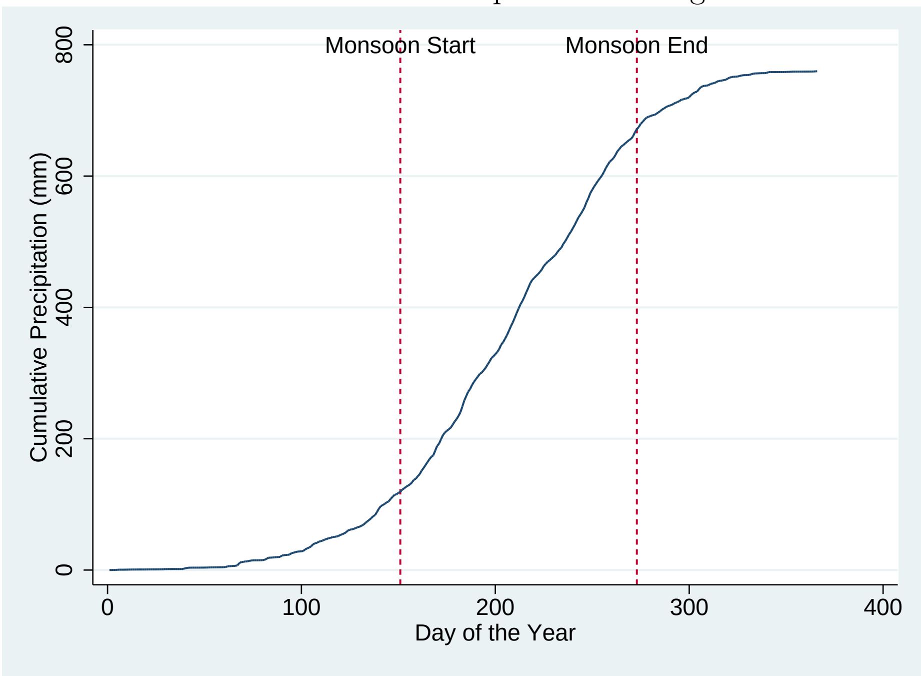

Figure 2. Kernel Density of Planting Time Relative to Onset Signal

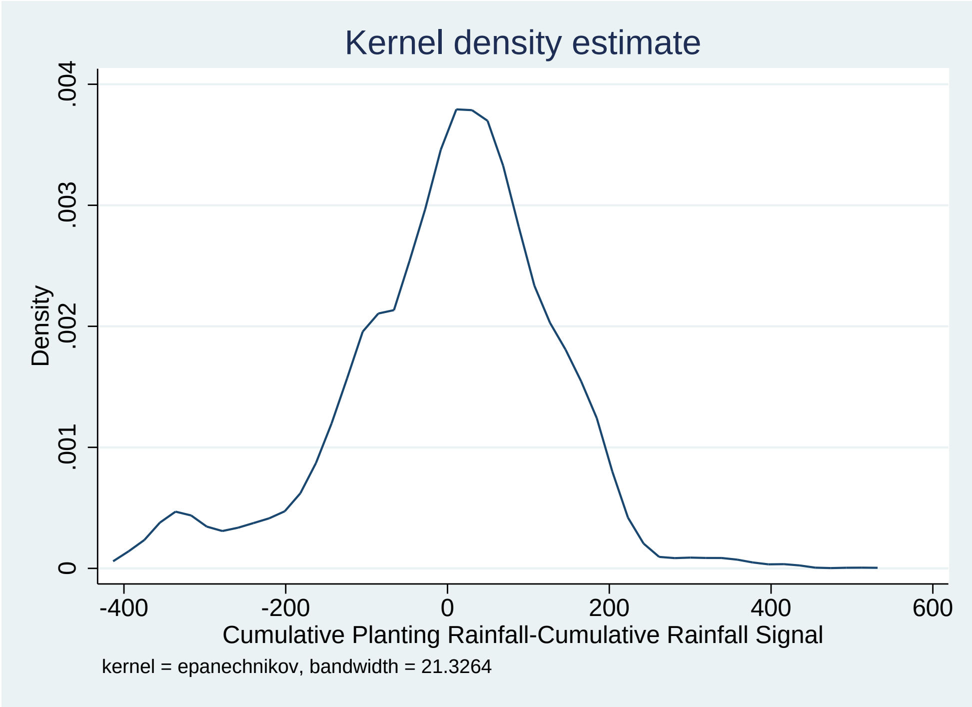

Figure 3. Relative Weights on Information: Varying Signal Variance

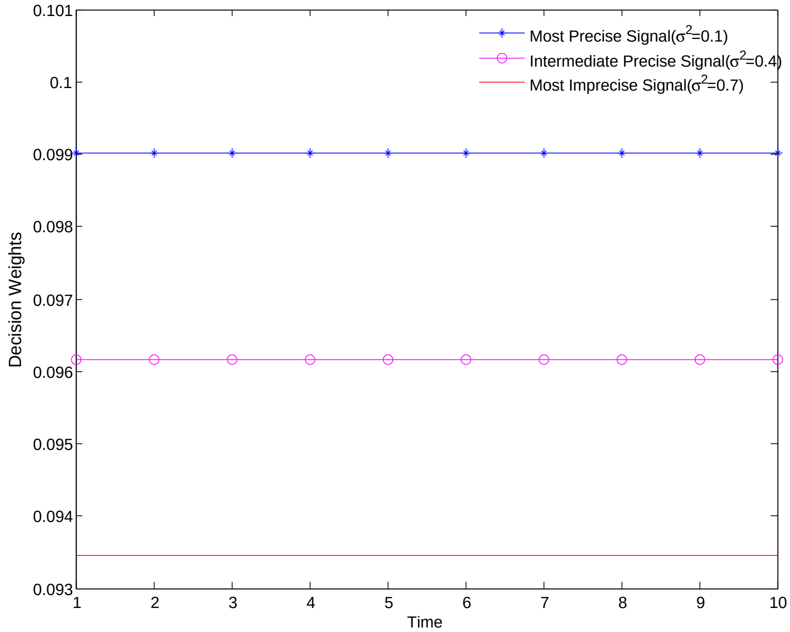

Figure 4. Relative Weights on Information: Varying State Variance

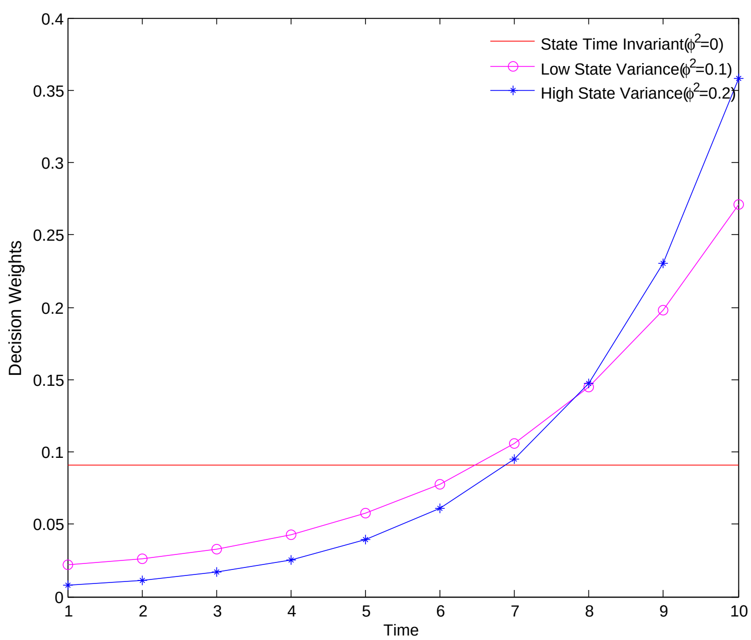

FIGURE 5. Relative Weights on Information:Varying Robustness

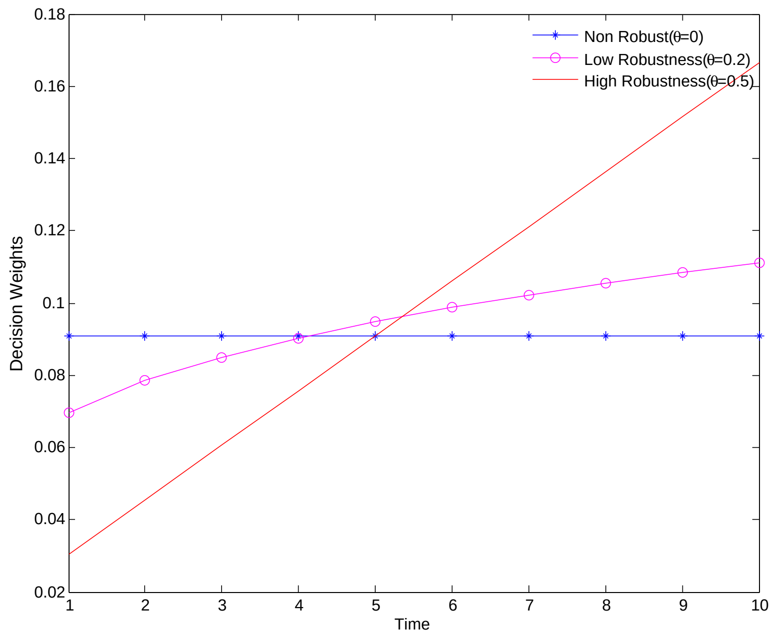

Figure 6. Rainfall Signal Timing for a High KLD Village and a Low KLD Village

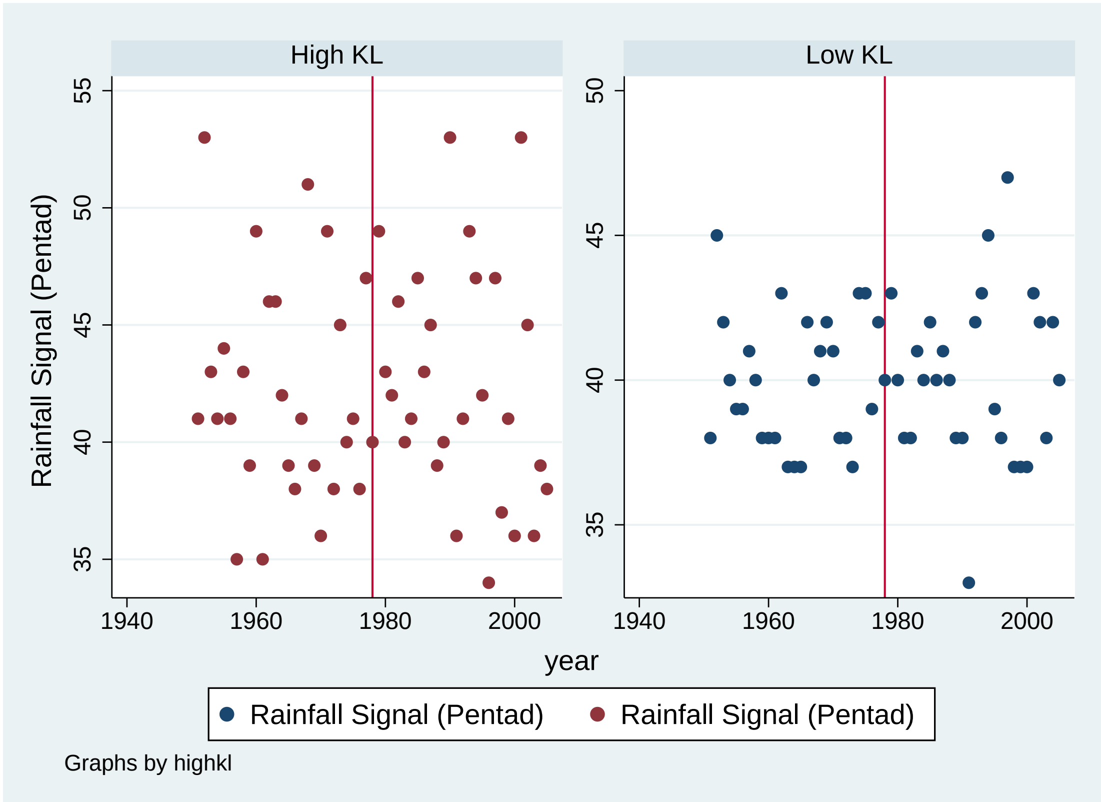

FIGURE 7. Maximum Forecast Errors by High and Low KLD Villages

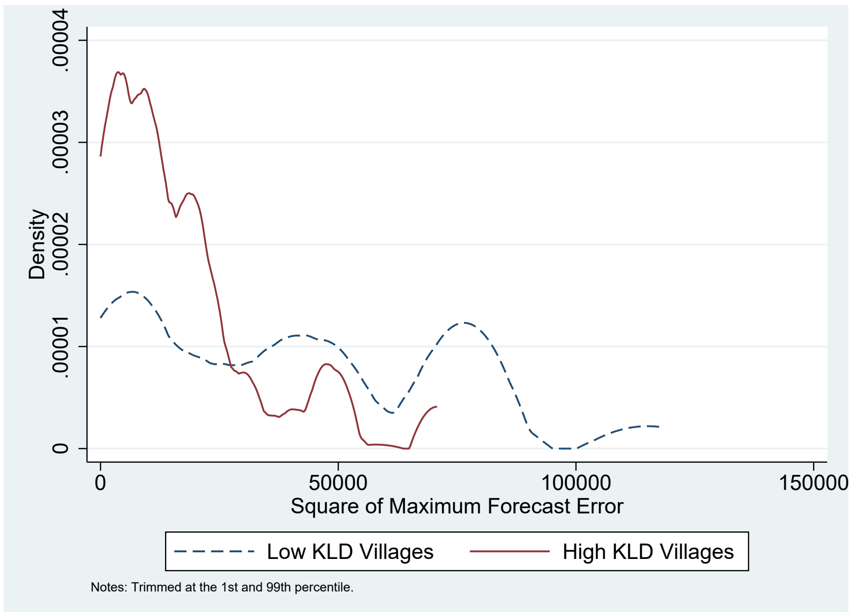

TABLE 1. Updating Rules

Model	*Updating Rules* Kalman Gain K_t	$Var(\widehat{\eta}_t - \eta_t)$ Σ_{t+1}
Normal Learning Model	$\frac{\Sigma_t}{\Sigma_t + \sigma^2}$	$\frac{\Sigma_t \sigma^2}{\Sigma_t + \sigma^2}$
Robust Normal Learning Model	$\frac{\Sigma_t}{\Sigma_t + \sigma^2(1 - \theta\Sigma_t)}$	$\frac{\Sigma_t \sigma^2}{\Sigma_t + \sigma^2(1 - \theta\Sigma_t)}$
Random Walk	$\frac{\Sigma_t}{\Sigma_t + \sigma^2}$	$\frac{\Sigma_t \sigma^2}{\Sigma_t + \sigma^2} + \phi^2$

TABLE 2. Estimated Parameters and Parameters Set to 0 in Each Learning Rule

Model	Signal Variance σ^2	*Parameters* State Variance ϕ^2	Ambiguity Aversion θ
Normal Learning Model	✓	0	0
Robust Normal Learning Model	✓	0	✓
Random Walk	✓	✓	0

Notes: Prior Variance (ω^2) normalized to 1.

Table 3. Summary Statistics

	Mean	SD
Rainfall Signal		
Profit Maximising Cumulative Rain Last Year	120.434	103.29
Plot-Level (Most Important Crop Planted)		
Cumulative Planting Rain	123.30	83.70
1(Irrigation)	0.51	0.49
Plot Area (Acres)	2.14	1.93
Irrigated Area (Acres)	0.40	1.04
Total Plot Value (Rs)	167221.70	203492.00
Profit (Rs)	9811.74	16478.98
Number of Household-Year Observations:	3,096	

The plot-year observations include data for plots that the household planted the most important crop earliest each year. The mean planting cumulative rainfall corresponds to 25th June-29th June.

Table 4. Impact of Rainfall Signal

	(1)	(2)	(3)	(4)
	Cumulative Planting Rainfall			
Profit-Maximizing Cumulative Rainfall Last Year (Contemporaneous Rainfall Signal)	0.134	0.226	0.0714	0.0884
	(0.0136)	(0.0157)	(0.0167)	(0.0226)
Fixed Effects	None	Year	Village, Year	Household, Year
Observations	2,551	2,551	2,551	2,551
Mean of Dependent Variable	123.84	123.84	123.84	123.84

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Standard errors are clustered at the household level. Profit-maximizing cumulative rainfall last year was the cumulative rainfall for the 5-day period planting in which maximized mean profits in the village last year. The sample includes the first day the household is observed to plant in a given year.

Table 5. Learning Models' Structural Parameters

	(1)	(2)	(3)
	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the rainfall signal ($\sigma 2$)	0.7938	1.1798	0.8269
	(0.1012)	(0.2291)	(0.1148)
Robustness Parameter (θ)		0.2438	
		(0.0661)	
Variance of optimal planting time ($\varphi 2$)			0.0184
			(0.0189)
Observations	2227	2227	2227
Akaike Information Criterion (AIC)	20305.52	20300.80	20305.79
Corrected Akaike Information Criterion (AICc)	20305.52	20300.81	20305.80
Bayesian Information Criterion (BIC)	20309.22	20310.22	20315.21

Notes: Robust standard errrors in parenthesis

Table 6. Learning Models' Structural Parameters by Kullback Leibler Divergence Criterion

Panel A: Villages with High Kullback Leibler Divergence Criterion

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the rainfall signal ($\sigma 2$)	0.7526	1.6013	0.8935
	(0.1454)	(0.3725)	(0.2272)
Robustness Parameter (θ)		0.4153	
		(0.0406)	
Variance of optimal planting time ($\varphi 2$)			0.2029
			(0.1011)
Observations	1097	1097	1097
Akaike Information Criterion (AIC)	7390.14	7386.11	7388.17
Corrected Akaike Information Criterion (AICc)	7392.83	7393.49	7395.56
Bayesian Information Criterion (BIC)	7392.83	7393.49	7395.56

Notes: Robust standard errrors in parenthesis

Panel B: Villages with Low Kullback Leibler Divergence Criterion

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the rainfall signal ($\sigma 2$)	0.7407	0.7899	0.7253
	(0.1360)	(0.2823)	(0.1366)
Robustness Parameter (θ)		0.0471	
		(0.2162)	
Variance of optimal planting time ($\varphi 2$)			-0.0074
			(0.0158)
Observations	1695	1695	1695
Akaike Information Criterion (AIC)	12089.16	12090.72	12091.75
Corrected Akaike Information Criterion (AICc)	12089.17	12090.74	12091.77
Bayesian Information Criterion (BIC)	12092.35	12099.10	12100.13

Notes: Robust standard errrors in parenthesis

Table 7. Learning Models' Structural Parameters by Irrigation Status

Panel A. Unirrigated Years

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the onset signal (σ2)	0.8339	1.6259	1.0115
	(0.2606)	(0.5471)	(0.3420)
Robustness Parameter (θ)		0.3694	
		(0.0859)	
Variance of optimal planting time (φ2)			0.1315
			(0.1128)
Observations	260	260	260
Akaike Information Criterion (AIC)	2328.76	2326.85	2327.84
Corrected Akaike Information Criterion (AICc)	2328.81	2326.94	2327.93
Bayesian Information Criterion (BIC)	2330.32	2331.97	2332.96

Notes: Robust standard errrors in parenthesis

Panel B. Irrigated Years

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the onset signal (σ2)	3.3836	4.9776	4.5792
	(0.8616)	(1.9284)	(1.7532)
Robustness Parameter (θ)		0.1692	
		(0.0878)	
Variance of optimal planting time (φ2)			0.2177
			(0.2260)
Observations	373	373	373
Akaike Information Criterion (AIC)	3347.76	3348.75	3348.95
Corrected Akaike Information Criterion (AICc)	3347.79	3348.82	3349.01
Bayesian Information Criterion (BIC)	3349.68	3354.60	3354.79

Notes: Robust standard errrors in parenthesis

Table 8. Learning Models' Structural Parameters: Crop Specific Rainfall Signal

	(1)	(2)	(3)
	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the onset signal ($\sigma 2$)	1.9174	2.8641	2.2677
	(0.2668)	(0.6029)	(0.3836)
Robustness Parameter (θ)		0.2261	
		(0.0669)	
Variance of optimal planting time ($\varphi 2$)			0.1596
			(0.1019)
Observations	1105	1105	1105
Akaike Information Criterion (AIC)	10068.56	10065.88	10066.49
Corrected Akaike Information Criterion (AICc)	10068.57	10065.91	10066.51
Bayesian Information Criterion (BIC)	10071.56	10073.90	10074.51

Notes: Robust standard errrors in parenthesis

Table 9. Maximum and Minimum Forecast Errors by Kullback Leibler Divergence Criterion

	(1)	(2)	(3)	(4)
	Maximum Squared Forecast Error		Minimum Squared Forecast Error	
1(High KLD Village)	-17,844	-3,719	124.5	524.9
	(8,452)	(1,396)	(208.6)	(136.9)
Additonal Controls	None	Mean Forecast Error	None	Mean Forecast Error
Observations	678	672	686	668
R-Squared	0.110	0.804	0.002	0.154
Mean of Dependent Variable	24616	24053	529.4	514.8

Notes: Robust standard errrors clustered at the village level in parenthesis

ONLINE APPENDIX

FIGURE A1. Location of Villages

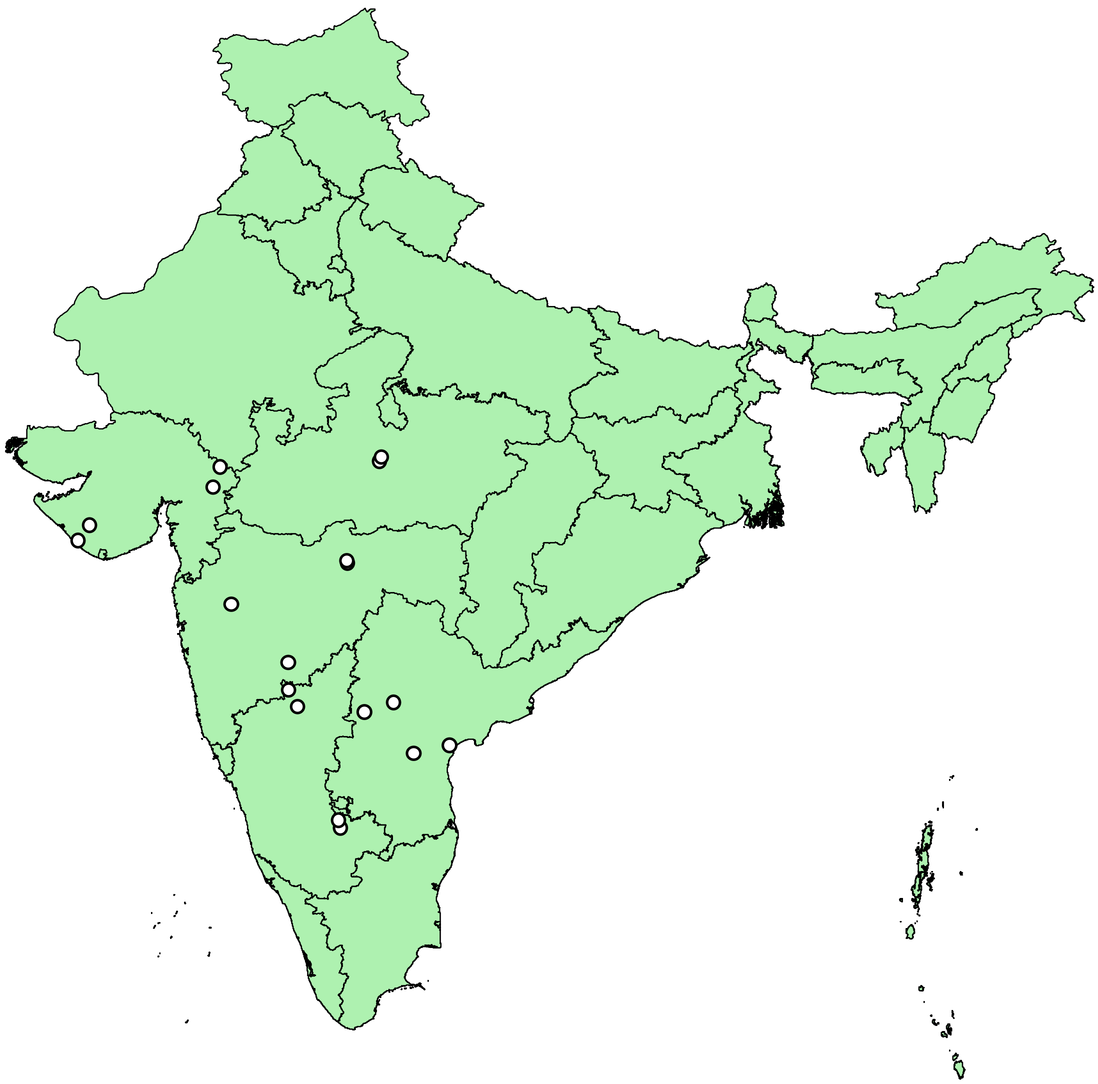

Table A1. Learning Models' Structural Parameters By Wealth

Panel A: Below Median Land Value Farmers

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the onset signal (σ2)	0.5474	1.3019	0.6591
	(0.1041)	(0.2015)	(0.1288)
Robustness Parameter (θ)		0.4912	
		(0.0433)	
Variance of optimal planting time (φ2)			0.1192
			(0.0479)
Observations	948	948	948
Akaike Information Criterion (AIC)	8697.40	8677.48	8683.49
Corrected Akaike Information Criterion (AICc)	8697.41	8677.51	8683.52
Bayesian Information Criterion (BIC)	8700.26	8685.19	8691.20

Notes: Robust standard errrors in parenthesis

Panel B: Above Median Land Value Farmers

	Normal Learning Model	Robust Normal Learning Model	Random Walk Kalman Filter
Variance of the onset signal (σ2)	0.7645	0.6326	0.6796
	(0.0826)	(0.2440)	(0.0700)
Robustness Parameter (θ)		-0.1781	-0.0504
		(0.3714)	(0.0067)
Variance of optimal planting time (φ2)			
Observations	1311	1311	1311
Akaike Information Criterion (AIC)	11784.71	-	-
Corrected Akaike Information Criterion (AICc)	11784.72	-	-
Bayesian Information Criterion (BIC)	11787.89	-	-

Notes: Robust standard errrors in parenthesis

Table A2. Impact of Deviation from Optimal Planting Time on Agricultural Profits

	(1)	(2)	(3)	(4)
	Agricultural Profit			
Log(Squared Deviation from Optimal Planting Rainfall)	-0.201	-0.135	-0.241	-0.208
	(0.0873)	(0.0737)	(0.0327)	(0.0511)
Fixed Effects	None	Village, Year	Village X Year	Household, Village X Year
Observations	3,083	3,083	3,083	3,083
R-Squared	0.053	0.170	0.337	0.546
Mean of Dependent Variable	6.00	6.00	6.00	6.00

Notes: Standard errors at the village-level in parentheses. The dependent variable is the inverse hyperbolic sine of agricultural profits. Log(Squared Deviation from Optimal Planting Rainfall) is the log of the squared deviation between cumulative rainfall in which the household planted and the optimal cumulative rainfall in the village that year

Table A3. Planting Date Choice as a Function of Prior Belief and Previous Year's Signals

	(1)	(2)	(3)	(4)	(5)
	Cumulative Planting Rainfall				
Prior (Planting Date in First Year)	0.379	0.264	0.0399	0.548	0.512
	(0.137)	(0.215)	(0.188)	(0.176)	(0.184)
Profit-Maximizing Cumulative Rainfall Last Year	0.562	0.541	0.278	0.243	0.254
	(0.155)	(0.130)	(0.0771)	(0.258)	(0.217)
Profit-Maximizing Cumulative Rainfall Two Years Ago		0.151	0.246	0.28	0.222
		(0.170)	(0.125)	(0.0362)	(0.174)
Profit-Maximizing Cumulative Rainfall Three Years Ago			0.215	0.127	0.196
			(0.213)	(0.137)	(0.0887)
Profit-Maximizing Cumulative Rainfall Four Years Ago				-0.00879	0.195
				(0.119)	(0.0910)
Profit-Maximizing Cumulative Rainfall Five Years Ago					-0.303
					(0.198)
Year of Data	3	4	5	6	7
Observations	582	545	452	199	161
R-squared	0.616	0.537	0.723	0.823	0.753

Notes: Robust standard errrors clustered at the village level in parenthesis

Appendix A. Farmer Learns from Multiple Signals, of Which Econometrician Observes Only One

Consider the case where the economic agent/the farmer has access to two signals, call them y_1 and y_2, to form his best estimate of the underlying state η, but the econometrician in trying to estimate the farmer's forecast (or more to the point, estimate unbiased decision weights assigned to signals in a given learning model) has access to only one of the signals, y_1.

In the econometrician's case of a single error-ridden signal y_1 generated according to

$$y_1 = \eta + \varepsilon_1 \tag{14}$$

with the distributional assumptions $\eta \sim N(\mu, \omega^2)$ and $\varepsilon_1 \sim N(0, \sigma^2)$ known to the farmer. Thus, the optimal predictor/estimate formed by the econometrician's estimating the farmers prediction will be:

$$\widehat{\eta}_1 \equiv E\left[\eta \mid y_1\right] = (1-\lambda)\,\mu + \lambda y_1 \tag{15}$$

where

$$\lambda \equiv \frac{\omega^2}{\sigma^2+\omega^2} = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2}+\frac{1}{\omega^2}} \equiv \frac{h_{\sigma^2}}{h_{\sigma^2}+h_{\omega^2}} \tag{16}$$

Assume that the second signal y_2 is generated by

$$y_2 = \eta + \varepsilon_2 \tag{17}$$

with the distributional assumption $\varepsilon_2 \sim N(0, \delta^2)$. As with Hyslop and Imbens (2001), the easiest assumption is to assume that the noise terms are uncorrelated, $Cov\left(\varepsilon_2, \varepsilon_2\right) = 0$ (This is relaxed later). In this case, the farmer's true estimate is

$$\widehat{\eta}_{12} \equiv E\left[\eta \mid y_1, y_2\right] = (1-\lambda_1-\lambda_2)\,\mu + \lambda_1 y_1 + \lambda_2 y_2$$

where $\lambda_1 \equiv \frac{h_{\sigma^2}}{h_{\sigma^2}+h_{\delta^2}+h_{\omega^2}}$ and $\lambda_2 \equiv \frac{h_{\delta^2}}{h_{\sigma^2}+h_{\delta^2}+h_{\omega^2}}$. The econometrician's estimate of the farmer's estimate can thus be thought of as an error-ridden measure of the truth i.e. $\widehat{\eta}_1 \equiv \widehat{\eta}_{12} + \zeta$ where the "measurement error" ζ is just by definition $\zeta \equiv \widehat{\eta}_1 - \widehat{\eta}_{12}$

So if we look at the correlation of the measurement error ζ with the "true" (or correct) estimator of the farmer, $\widehat{\eta}_{12}$, the covariance is:

$$Cov(\widehat{\eta}_{12}, \zeta) = Cov(\widehat{\eta}_{12}, \widehat{\eta}_1) - Var(\widehat{\eta}_{12}) \tag{18}$$

Intuitively, even from this expression alone, we can see that as the second signal y_2 adds little to the farmer's forecast, then $Cov(\widehat{\eta}_{12}, \widehat{\eta}_1)$ approaches $Var(\widehat{\eta}_{12})$ in the limit, and this measurement error is zero.

We can show that

$$Cov(\widehat{\eta}_{12}, \widehat{\eta}_1) = -\lambda\lambda_2\sigma^2 \tag{19}$$

Using this, we can show that the discrepancy between the "incomplete" estimate of the farmer's forecast given by the econometrician's $\widehat{\eta}_1$ is uncorrelated with the farmer's *estimate*. The implication of showing that the the discrepancy between the "incomplete" estimate of the farmer's forecast given by the econometrician's $\widehat{\eta}_1$ is uncorrelated with the farmer's *estimate* is in contrast to the usual measurement error problem, where the observed variable is (under the so-called Classical Measurement Error assumptions) correlated with the measurement error, and it is this property that creates bias in OLS. In this case, because the observed/constructed variable used by the econometrician is uncorrelated with the error as regards the true measure $\widehat{\eta}_{12}$, then in an OLS regression, for example, no measurement error problem would manifest itself.

By definition of the measurement error, we have:

$$Cov(\widehat{\eta}_1, \zeta) = Var(\widehat{\eta}_1) - Cov(\widehat{\eta}_{12}, \widehat{\eta}_1)$$

The second term is computed to be

$$Cov(\widehat{\eta}_{12}, \widehat{\eta}_1) = \lambda\omega^2 = \frac{h_{\sigma^2}\omega^2}{h_{\sigma^2} + h_{\omega^2}}$$

The first term is

$$Var(\widehat{\eta}_1) \quad = \quad Var(\lambda y_1) = \lambda^2\left(\omega^2 + \sigma^2\right) = \lambda\omega^2$$

Thus,

$$Cov(\widehat{\eta}_1, \zeta) = \lambda\omega^2 - \lambda\omega^2 = 0$$

Thus, the measurement error induced by the econometrician using only a subset of the signals in her estimate of the farmer's forecast is actually uncorrelated with her estimate,

This result holds even allowing for a non-zero covariance in ε_1 and ε_2

The derivation of this more general case starts by considering the each signal net of its linear dependence on the other signal - i.e.

$$\xi_1 \equiv y_1 - \frac{Cov(y_1, y_2)}{Var(y_2)}y_2$$

and

$$\xi_2 \equiv y_2 - \frac{Cov(y_1, y_2)}{Var(y_1)}y_1$$

As these expressions exhibit obvious symmetry, we can derive results for one, then extrapolate to the other by the symmetry of the expressions. In addition,

$$\lambda_1 \equiv \frac{Cov(\eta, \xi_1)}{Var(\xi_1)}$$

by the Frisch-Waugh Theorem. In terms of the numerator of λ_1, we have that

$$Cov(\eta,\xi_1) = \omega^2 - \frac{Cov\left(y_1,y_2\right)}{Var\left(y_2\right)}\omega^2 = \frac{\omega^2}{Var\left(y_2\right)}\left[Var\left(y_2\right) - Cov\left(y_1,y_2\right)\right]$$

and for the denominator, we have that

$$Var\left(\xi_1\right) = \frac{1}{Var\left(y_2\right)}\left\{Var\left(y_1\right)Var\left(y_2\right) - \left[Cov\left(y_1,y_2\right)\right]^2\right\}$$

So that in taking the quotient of these two expressions, and dividing out the common denominator of $Var\left(y_2\right)$, we thus have that

$$\lambda_1 = \frac{\omega^2\left[Var\left(y_2\right) - Cov\left(y_1,y_2\right)\right]}{Var\left(y_1\right)Var\left(y_2\right) - \left[Cov\left(y_1,y_2\right)\right]^2}$$

and thus by the symmetry of the problem, we can thus also immediately state that

$$\lambda_2 = \frac{\omega^2\left[Var\left(y_1\right) - Cov\left(y_1,y_2\right)\right]}{Var\left(y_1\right)Var\left(y_2\right) - \left[Cov\left(y_1,y_2\right)\right]^2}$$

Allowing for the more general case that $Cov\left(y_1,y_2\right) \neq \omega^2$,

$$\begin{aligned}
Cov(\widehat{\eta}_{12},\widehat{\eta}_1) &= Cov(\lambda_1 y_1 + \lambda_2 y_2, \lambda y_1) \\
&= \lambda\lambda_1 Var(y_1) + \lambda\lambda_2 Cov\left(y_1,y_2\right) \\
&= \frac{\lambda\omega^2\left\{\left[Var\left(y_2\right) - Cov\left(y_1,y_2\right)\right]Var(y_1) + \left[Var\left(y_1\right) - Cov\left(y_1,y_2\right)\right]Cov\left(y_1,y_2\right)\right\}}{Var\left(y_1\right)Var\left(y_2\right) - \left[Cov\left(y_1,y_2\right)\right]^2} = \lambda\omega^2
\end{aligned}$$

Thus, we can conclude that

$$Cov(\widehat{\eta}_1,\zeta) = Var(\widehat{\eta}_1) - Cov(\widehat{\eta}_{12},\widehat{\eta}_1) = \lambda\omega^2 - \lambda\omega^2 = 0$$

holds even when the signals used by the farmer have a common noise component - i.e. the econometrician's estimator of the farmer's forecast/estimate remains uncorrelated with the measurement error$\zeta \equiv \widehat{\eta}_1 - \widehat{\eta}_{12}$.

Appendix B. Equivalent Versions of the Robust Learning Model

Recall that the robust learning was captured by the following equation

$$\widehat{\eta}_{t+1} = \left(1 - K_t^R\right)\widehat{\eta}_t + K_t^R y_t,$$

where

$$K_t^R = \frac{\Sigma_t^R}{\Sigma_t^R + \sigma^2(1-\theta\Sigma_t^R)}, \qquad \Sigma_t^R = \frac{\Sigma_{t-1}^R}{\Sigma_{t-1}^R + \sigma^2(1-\theta\Sigma_{t-1}^R)}, \qquad \Sigma_0^R = \omega^2.$$

In this appendix, I provide two alternate versions of the robust learning model which lead to the same updating equation. This serves the following purpose: (i) it highlights the generality of the framework, (ii) provides intuition on the derivation of the above equation and (iii) demonstrates the role played by the particular functional form assumptions. Throughout this section, I assume that the underlying state is time invariant ($\phi^2 = 0$) which is the model I estimate.

In order to reduce notation, I eliminate the constants a_t from the farmer's maximization problem and express his max-min problem

$$\underset{\widehat{\eta}_t}{\operatorname{argmax}} \min_{G_t} \left\{ a_t - \int (\widehat{\eta}_t - \eta_t)^2 \, dG_t(\eta_t) + \sum_{t'=1}^{t-1} \left[a_{t'} - \int (\widehat{\eta}_{t'} - \eta_t)^2 \, dG_t(\eta_t) \right] + \frac{1}{\theta} KL(G_t, F_t) \right\}$$

as the following equivalent min-max problem instead

$$\underset{\widehat{\eta}_t}{\operatorname{argmin}} \max_{G_t} \left\{ \int (\widehat{\eta}_t - \eta_t)^2 \, dG_t(\eta_t) + \sum_{t'=1}^{t-1} \left[\int (\widehat{\eta}_{t'} - \eta_t)^2 \, dG_t(\eta_t) \right] - \frac{1}{\theta} KL(G_t, F_t) \right\}.$$

The following proposition is a summary of results that can be found in the literature. I first discuss its implication and then provide a proof (for the ease of reference for the interested reader).

Proposition 1. *The following optimization problems are minimized by the same unique* $\widehat{\eta}_t$*:*

(A)

$$\underset{\widehat{\eta}_t}{\operatorname{argmin}} \max_{G_t} \left\{ \int (\widehat{\eta}_t - \eta)^2 \, dG_t(\eta) + \sum_{t'=1}^{t-1} \left(\int (\widehat{\eta}_{t'} - \eta)^2 \, dG_t(\eta) \right) - \frac{1}{\theta} KL(G_t, F_t) \right\}$$

(B)

$$\underset{\widehat{\eta}_t}{\operatorname{argmin}} \max_{\eta} \left\{ (\widehat{\eta}_t - \eta)^2 + \sum_{t'=1}^{t-1} (\widehat{\eta}_{t'} - \eta)^2 - \frac{1}{2\theta} \left[\frac{(\eta - \mu_0)^2}{w^2} + \sum_{t'=1}^{t-1} \frac{(\eta - y_{t'})^2}{\sigma^2} \right] \right\}.$$

(C)

$$\underset{\widehat{\eta}_t}{\operatorname{argmin}} \left\{ \frac{1}{\theta} \log \mathbb{E}\left(e^{\theta L_t}\right) \right\}, \quad \textit{where} \ \ L_t := (\widehat{\eta}_t - \eta)^2 + \sum_{t'=1}^{t-1} (\widehat{\eta}_{t'} - \eta)^2.$$

The objective function in (B) is the certainty equivalent version of the robust profit maximization problem (A) which was presented and discussed in Section 3. Note that in this problem, nature is not choosing a distribution G_t of the optimal planting time but is instead choosing an actual planting time η itself to minimize the cumulative profits (captured by the first two terms). As in (A), the problem is disciplined by the last penalty term (in the square brackets); intuitively, nature receives a bigger penalty from picking a planting time which makes the prior mean and the subsequent signals more "misleading."

The derivation of the simple linear updating equation (9) is immediate from this restatement of the farmer's optimization problem. Observe that, the objective function is quadratic which implies that the optimal planting time η picked by nature (for any choice $\widehat{\eta}_t$ by the farmer) will be a linear function of chosen planting dates, the prior and the observed signals. In turn, this implies the optimal choice of $\widehat{\eta}_t$ by the farmer will be a linear function of the prior and the observed signals.

Finally, it is worth pointing out that there is also a purely Bayesian formulation given by (C) (see Whittle (1981)) which yields the same updating equation (9). This formulation is more appropriate for a learning environment where a statistician is trying learn an underlying parameter by minimizing the cumulative loss given by $\frac{1}{\theta}\log\mathbb{E}\left(e^{\theta L_t}\right)$. In this context, the choice of loss function requires no economic interpretation per se. However, in my setting, farmers choose planting dates to maximize period t profits and, hence, there is no obvious compelling interpretation for an objective function such as that in part (C) which includes the expected profits in all periods prior to t as well.

The following is the proof of the above proposition.

Proof of Proposition 1: I first show that in problems (A) and (B), the optimally chosen $\widehat{\eta}_t$ is the same.

First observe that the objective function in (A) can be rewritten as follows:

$$\begin{aligned}
&\int\left(\widehat{\eta}_t-\eta\right)^2 dG_t(\eta)+\sum_{t'=1}^{t-1}\int\left(\widehat{\eta}_{t'}-\eta\right)^2 dG_t(\eta)-\frac{1}{\theta}KL(G_t,F_t)\\
&=\int\left(\widehat{\eta}_t-m_1(G_t)+m_1(G_t)-\eta\right)^2 dG_t(\eta)+\sum_{t'=1}^{t-1}\int\left(\widehat{\eta}_{t'}-m_1(G_t)+m_1(G_t)-\eta\right)^2 dG_t(\eta)\\
&\quad-\frac{1}{2\theta}KL(G_t,F_t)\\
&=\left(\widehat{\eta}_t-m_1(G_t)\right)^2+m_2(G_t)+\sum_{t'=1}^{t-1}\left[\left(\widehat{\eta}_{t'}-m_1(G_t)\right)^2+m_2(G_t)\right]-\frac{1}{\theta}KL(G_t,F_t)
\end{aligned}\tag{20}$$

where $m_1(\cdot)$ and $m_2(\cdot)$ (the first and second moments) are used to denote the mean and variance of a distribution (in this case G_t).

Recall, that the distribution F_t in the above expression is obtained from Bayesian updating (as in the standard NLM) and is normal with mean and variance given by

$$m_1(F_t)=\frac{\sum_{t'=1}^{t-1}\frac{y_{t'}}{\sigma^2}+\frac{\mu_0}{w^2}}{\frac{t-1}{\sigma^2}+\frac{1}{w^2}}\text{ and }m_2(F_t)=\frac{1}{\frac{t-1}{\sigma^2}+\frac{1}{w^2}}.\tag{21}$$

A well known result in information theory is that for a fixed variance, the normal distribution has the highest entropy. A simple to show consequence of this result is that when nature maximizes the above objective function (20) for any choice $\widehat{\eta}_t$ by the farmer, she will pick G_t to be a normal distribution as this will minimize the Kullback-Liebler (KL) divergence (for any mean and variance $m_1(G_{t+1})$ and $m_2(G_{t+1})$) from the normal distribution F_t.

The KL divergence $KL(G_t, F_t)$ between two normal distributions is given by

$$KL(G_t, F_t) = \frac{1}{2}\log\left(\frac{m_2(F_t)}{m_2(G_t)}\right) + \frac{m_2(G_t) + (m_1(G_t) - m_1(F_t))^2}{m_2(F_t)} - \frac{1}{2}.$$

Thus nature effectively just picks the mean and variance of G_t (as we have argued that it will be normal). This allows us to express the optimization problem (A) as

$$\begin{aligned}
&\operatorname*{argmin}_{\widehat{\eta}_t} \max_{m_1(G_t), m_2(G_t)} \Bigg\{ (\widehat{\eta}_t - m_1(G_t))^2 + m_2(G_t) + \sum_{t'=1}^{t-1}\left[(\widehat{\eta}_{t'} - m_1(G_t))^2 + m_2(G_t)\right] \\
&\qquad\qquad - \frac{1}{\theta}\left[\frac{1}{2}\log\left(\frac{m_2(F_t)}{m_2(G_t)}\right) + \frac{m_2(G_t) + (m_1(G_t) - m_1(F_t))^2}{2m_2(F_t)} - \frac{1}{2}\right]\Bigg\} \\
&= \operatorname*{argmin}_{\widehat{\eta}_t} \max_{m_1(G_t)} \Bigg\{ (\widehat{\eta}_t - m_1(G_t))^2 + \sum_{t'=1}^{t-1}(\widehat{\eta}_{t'} - m_1(G_t))^2 - \frac{1}{\theta}\left[\frac{(m_1(G_t) - m_1(F_t))^2}{2m_2(F_t)}\right]\Bigg\} \qquad (22)
\end{aligned}$$

where we can ignore nature's maximization with respect to $m_2(G_t)$ as these terms are additively separable and hence do not affect the farmer's choice of $\widehat{\eta}_t$. We change the variable name and replace $m_1(G_t)$ by η and plug in the expressions for $m_1(F_t)$ and $m_2(F_t)$ to restate (A) as

$$\operatorname*{argmin}_{\widehat{\eta}} \max_{\eta} \Bigg\{ (\widehat{\eta}_t - \eta)^2 + \sum_{t'=1}^{t-1}(\widehat{\eta}_{t'} - \eta)^2 - \frac{1}{2\theta}\left[\frac{(\sum_{t'=1}^{t-1}\frac{\eta - y_{t'}}{\sigma^2} + \frac{\eta-\mu_0}{w^2})^2}{\frac{t-1}{\sigma^2} + \frac{1}{w^2}}\right]\Bigg\}. \qquad (23)$$

Observe that both (23) and (B) are both optimization problems with quadratic objective functions that differ only in their last term. Observe also that the farmer's choice $\widehat{\eta}_t$ only enters the first term (which is identical across both problems). Thus, it suffices to show that the η chosen by nature which minimizes the objective is the same across both problems.

This last step is easily show by noting that the first order conditions of both problems (23) and (B) are identical and linear (since the objective function is quadratic, these are also sufficient) which implies that they will both have the same unique solution. To see this, note that the last terms (in the square brackets) are the only terms that differ in both the objective functions but that they have the same derivatives:

$$\begin{aligned}
&\frac{d}{d\eta}\left\{\frac{1}{2\theta}\left(\frac{(\eta-\mu_0)^2}{w^2} + \sum_{k=1}^{t}\frac{(\eta - y_t)^2}{\sigma^2}\right)\right\} = \frac{d}{d\eta}\left\{\frac{1}{2\theta}\left[\frac{(\sum_{t'=1}^{t-1}\frac{\eta - y_{t'}}{\sigma^2} + \frac{\eta-\mu_0}{w^2})^2}{\frac{t-1}{\sigma^2} + \frac{1}{w^2}}\right]\right\} \\
&= \frac{1}{\theta}\left(\frac{(\eta-\mu_0)}{w^2} + \sum_{t'=1}^{t-1}\frac{(\eta - y_{t'})}{\sigma^2}\right).
\end{aligned}$$

Hence, in both problems (A) and (B), the optimally chosen $\widehat{\eta}_t$ is the same.

I now show that in problems (A) and (C), the optimally chosen $\widehat{\eta}_t$ is the same.

Problem (C) is a minimization problem where the farmer minimizes $2\theta^{-1}\log\mathbb{E}(e^{\frac{1}{2}\theta L_t})$ where the expectation is taken with respect to the distribution F_t obtained from Bayesian updating (with mean and variance given in equation (21) above). Plugging in the distribution, this problem can be stated as

$$\operatorname*{argmin}_{\widehat{\eta}_t}\left\{\frac{1}{\theta}\int_{-\infty}^{\infty}\exp\left(\theta\left((\widehat{\eta}_t-\eta)^2+\sum_{t'=1}^{t-1}(\widehat{\eta}_{t'}-\eta)^2\right)\right)\exp\left(-\frac{\left(\eta-\frac{\sigma^2\mu_0+w^2\sum_{t'=1}^{t-1}y_{t'}}{\sigma^2+(t-1)w^2}\right)^2}{2\frac{\sigma^2w^2}{\sigma^2+(t-1)w^2}}\right)d\eta\right\}$$
$$=\operatorname*{argmin}_{\widehat{\eta}_t}\left\{\frac{1}{\theta}\int_{-\infty}^{\infty}\exp\left(\theta\left[(\widehat{\eta}_t-\eta)^2+\sum_{t'=1}^{t-1}(\widehat{\eta}_{t'}-\eta)^2-\frac{1}{2\theta}\left(\frac{(\sum_{t'=1}^{t-1}\frac{\eta-y_{t'}}{\sigma^2}+\frac{\eta-\mu_0}{w^2})^2}{\frac{t-1}{\sigma^2}+\frac{1}{w^2}}\right)\right]\right)d\eta\right\}.$$

I now invoke Lemma 1 (stated and proved below) which shows that the solution to the above problem is the same as the following min-max problem.

$$\operatorname*{argmin}_{\widehat{\eta}}\max_{\eta}\left\{(\widehat{\eta}_t-\eta)^2+\sum_{t'=1}^{t-1}(\widehat{\eta}_{t'}-\eta)^2-\frac{1}{2\theta}\left[\frac{(\sum_{t'=1}^{t-1}\frac{\eta-y_{t'}}{\sigma^2}+\frac{\eta-\mu_0}{w^2})^2}{\frac{t-1}{\sigma^2}+\frac{1}{w^2}}\right]\right\}.$$

Observe that this problem is identical to (23) which shows the required equivalence to the problem (A).

I end the proof with the lemma invoked above.

Lemma 1. *Let*

$$x^*\in\operatorname*{argmin}_{x}\int_{-\infty}^{\infty}\exp\left(-ay^2+2byx+2cy+ex^2+fx\right)dy,$$

where it is assumed that the solution exists. Then it follows that

$$x^*\in\operatorname*{argmin}_{x}\max_{y}\left\{-ay^2+2byx+2cy+ex^2+fx\right\}.$$

Proof. We prove this lemma by separately solving each of the two minimization problems. We first consider

$$\min_x\int_{-\infty}^{\infty}\exp\left(-ay^2+2byx+2cy+ex^2+fx\right)dy$$

Integrating with respect to y, we can restate this problem as

$$\min_x\sqrt{\frac{\pi}{a}}\exp\left(\frac{(2bx+2c)^2}{4a}+ex^2+fx\right).$$

It is equivalent to minimize the term in the brackets, or equivalently, to solve

$$\min_x\left(\frac{(bx+c)^2}{a}+ex^2+fx\right).$$

Taking a first order condition (which is also sufficient for a quadratic objective function), we get

$$\frac{2b(bx^*+c)}{a}+2ex^*+f=0\implies x^*=-\frac{2bc+af}{2(b^2+ae)}.$$

We now solve,

$$\max_y \left\{-ay^2 + 2byx + 2cy + ex^2 + fx\right\},$$

for a fixed x. Taking a first order condition with respect to y, we get

$$-2ay^*(x) + 2bx + 2c = 0 \implies y^*(x) = \frac{bx+c}{a}.$$

Plugging into the minimization problem, we can rewrite it as

$$\min_x \left\{-a\left(\frac{bx+c}{a}\right)^2 + 2b\frac{bx^2+cx}{a} + 2c\frac{bx+c}{a} + ex^2 + fx\right\}.$$

Taking a first order condition, we get

$$\begin{aligned} & -\frac{2b(bx^*+c)}{a} + 2b\frac{2bx^*+c}{a} + \frac{2bc}{a} + 2ex^* + f = 0 \\ \implies & -2b^2x^* + 4b^2x^* + 2aex^* = -(af + 2bc) \\ \implies & x^* = -\frac{2bc+af}{2(b^2+ae)}, \end{aligned}$$

which completes the proof. □

□