Learning from Weather Forecasts and Short-Run Adaptation: Evidence from an At-Scale Experiment

Jess Rudder and Davide Viviano *

October 1, 2024

Abstract

Weather forecasts are an example of a public good that is often distributed at-scale without charging users. This feature can make measuring the benefits of weather forecast distribution challenging because information spillovers are likely; people like to talk about the weather and weather information is often available from a variety of sources. Despite the ubiquity of weather information, small-scale farmers often lack access to high-quality weather forecasts that are tailored to help them make production decisions. We implement a randomized experiment with 400,000 cotton growers in Pakistan and vary the share of farmers treated with large clusters (tehsils). We show that treated and untreated farmers in high-saturation clusters update their farming behavior in line with forecasts. Directly treated farmers in high saturation tehsils are 37-67% more likely to avoid rain when irrigating and applying fertilizer and pesticides. Control farmers in highly saturated tehsils are 22-46% more likely to avoid rain compared to controls in low-saturation tehsils. For heat avoidance, results follow a similar pattern but are statistically weaker. Direct information sharing is a plausible pathway - control farmers in high saturation areas were 8% more likely than control farmers in low saturation areas to report discussing weather information with peers. At the end of the season, estimates for yields are positive but imprecise and there is evidence that input expenditure increased.

*Emails: jrudder@uchicago.edu and dviviano@fas.harvard.edu. We thank Chris Barrett, Diego Cardoso, Michael Carter, Jagori Chatterjee, Josh Deutschmann, Brian Dillon, Todd Gerarden, Alan Griffith, Tomoko Harigaya, Rachel Heath, Andrew Hultgren, Maulik Jagnani, Sarah Janzen, Khushbakht Jamal, Alex Lehner, Travis Lybbert, Hope Michelson, Supreet Kaur, Michael Kremer, Amir Jina, Arman Rezaee, Adeel Shafqat, Giulio Schinaia, Jessica Zhu and seminar participants at University of California Davis, University of Illinois, University of Washington, Cornell University, IPA/Global Poverty Research Lab Annual Researcher Gathering, Advances in Field Experiments at the University of Chicago, MeasureDev Conference at the World Bank for helpful comments and suggestions.

1 Introduction

It can be challenging to measure the benefits of large-scale information campaigns. Low delivery costs can justify wide distribution without spending additional resources to target participants with specific characteristics, even when information take-up varies. Once learned, some types of information are easy to share with neighbors, implying that spillovers are likely to attenuate treatment effects measurement. Weather forecasts are an example of an information public good that is often distributed widely at no cost to recipients with large public benefits (Rosenzweig and Udry, 2019; Molina and Rudik, 2023; Shrader et al., 2023). Yet, in many low- and middle-income countries, forecasts fail to reach small-scale farmers who would benefit from weather information when making production and investment decisions (Linsenmeier and Shrader, 2023).

Agricultural production is inherently risky due to uninsured weather shocks that threaten the livelihoods of small-scale farmers around the world (Lesk et al., 2016; Wollburg et al., 2024). Weather forecasts are an important tool to help farmers learn and adapt to new weather patterns caused by climate change. Short-range weather forecasts can help farmers plan immediate agricultural activities and take preventative measures that lower the likelihood of losses due to unforeseen weather. Although short-term weather services are widely used by farmers around the world, it is not clear whether short-range weather forecasts provide actionable information for farmers.

We tested an optimal design strategy to detect treatment and spillover effects in a phone-based advisory service that reaches 400,000 farmers in Pakistan. Our research design accounts for spillovers by assigning clusters (tehsils) to have variation in treatment intensity of the share of users per cluster assigned to any treatment. We first randomly assign tehsils to saturation levels based on the share of phone numbers available in each tehsil. We then randomly assign phone numbers to the control arm where they receive standard agronomic advisory messages throughout the growing season. Or, they are assigned to treatment where in addition to adivsory messages, they also receive localized weather forecasts. In order to measure weather beliefs and behaviors throughout the season, we collected high-frequency phone surveys with 400 farmers per week to learn how short-range weather forecasts affect: 1) input application timing (fertilizer, pesticides, irrigation), 2) information sharing with peer farmers in highly treated areas, and 3) agricultural productivity.

One feature of our setting is that the experimental frame includes all 400,000 phone numbers in 35 tehsils. The phone number lists were provided by the government partner to the NGO implementer. The most granular location information available was at the tehsil level, roughly equivalent to a county in the US. The NGO sent automated voice messages of weather forecasts with 2-day lead time that contained information about the expected rain, maximum temperature, and minimum temperature. A more ideal set-up for information interventions would be to first identify smaller cluster units - such as a household, neighborhood, or village to increase the probability that participants know each other. However, with administrative data and government partners, this type of recruitment is often infeasible and raises start-up costs.

Another feature of our setting is that there is a high potential for interference and non-compliance. Many farmers in the control group had the option to seek weather information from other sources - and many did - nearly 50% of control farmers reported receiving weather information in the prior week, compared to 60% in the treatment groups.[1] As anywhere in the world, it is also common to talk about the weather. In pilot surveys, nearly 80% of farmers reported discussing weather information with peers in the prior week. In this information environment, interference (information sharing between treatment and control) and non-compliance (control accessing weather information and treated declining take-up) can decrease power to estimate effects and highlights the importance of varying cluster saturation in order to identify treatment effects. The saturation design allows us to separately identify intention to treat effects, total spillover effects, spillovers on the treated in high saturation tehsils, and spillovers on the untreated in high saturation tehsils.

We focus on four categories of outcomes: information recall, sharing, farming behavior, and productivity. For forecast recall, the total spillover effect exhibits a positive and significant relationship, a 10% increase in the tehsil share treated increases the probability that farmer predictions align with forecasts by 1.02 percentage points. Results are similar for predictions of maximum and minimum temperature - directly treated farmers have more aligned predictions and the spillover effect is larger in magnitude. When decomposing spillovers between untreated and treated farmers, we see that treated farmers in highly saturated tehsils are twice as likely to have rainfall predictions that align with forecasts compared to treated farmers in low saturation tehsils, although the difference is not statistically significant.

Treatment induces information sharing by treated farmers in both low and high saturation tehsils - they are both 29-34% more likely to share information compared to control farmers. However, compared to control farmers in low saturation tehsils, control farmers in high saturation

[1] Although farmers may access weather forecasts from other sources, these forecasts were less granular and had different lead times than those provided by the program.

tehsils were 13% more likely to *receive* weather information from peers, suggesting that direct information sharing is a plausible channel that drives differences in behavior.

For farming behavior, we measured whether farmers carried out farming tasks, such as irrigation, fertilizing, pesticides, and other inputs on days without rain. Avoiding rain is a potentially important short-run adaptation because it increases resource efficiency and reduces runoff from agricultural inputs. Directly treated farmers in high saturation tehsils are 40-68% more likely to avoid rain when irrigating and applying fertilizer and pesticides. Control farmers in highly saturated tehsils are 24-47% more likely to avoid rain compared to controls in low-saturation tehsils. Treated farmers in low saturation tehsils are 3-14% more likely to avoid rain, but differences are less precise. Patterns are similar for avoiding heat. Heat avoidance is less directly related to input efficiency, but improves welfare by reducing exposure to high heat, which can have adverse health effects. Helping agricultural workers avoid extreme heat is an important policy goal.

For agricultural productivity, we see marginally positive increase for treated in low saturation tehsils, and positive but insignificant effects on yields in high saturation tehsils. We also see negative but insignificant effects on profits in high saturation tehsils and mixed results on revenue. It is possible that ambiguous profit effects are driven by the fact that treated and spillover farmers have higher input expenditure. There is also some evidence that they harvested later and earned lower output prices, although estimates are noisy. One explanation is that exercising flexibility in the timing of inputs caused farmers to adjust spending in a way that was costly in the short-run. Taken together, these findings suggest that short-run forecasts can provide meaningful information to farmers that help them increase efficiency by applying inputs on days when it does not rain and that information spillovers are meaningful.

More accurate forecasts increased take-up by recipients, using our preferred measures that focus on non-zero rainfall forecasts using both farmer-reported and satellite measures. On average across clusters, rain forecasts were accurate for 79% of days, and the worst-performing tehsils were accurate 61% of days. Among participants in the 2022 phone survey, 75% listened to at least one weather message and 77% listened to at least one advisory message, but average listening rates were closer to 30% indicating that there was high variation in engagement. A 10% increase in the share of days with an accurate rainfall forecast leads to a 1.6-2.9% increase in average take-up.

Consistent with the agricultural development literature on social learning and spillovers, prior studies among small-scale farmers suggest that information sharing is pervasive (Foster and Rosenzweig, 1995; Munshi, 2004; Bandiera and Rasul, 2006). In the context of learning from forecasts, previous studies showed that participants learn to interpret forecasts with relatively little training (Barnett-Howell, 2021; Ahmad et al., 2022; Cole et al., 2023). We contribute to a growing literature that evaluates the efficacy and cost effectiveness of delivering weather forecasts to farmers (Fafchamps and Minten, 2012;Fosu et al., 2018; Camacho and Conover, 2019; Yegbemey et al., 2023). Recently, Burlig et al. (2024) show that Indian farmers with access to better monsoon onset forecasts increased agricultural investment. Rosenzweig and Udry (2019) establish that accurate forecasts are particularly valuable and that Indian farmers learn accuracy overtime and make productive investment decisions accordingly. Digital agriculture services can increase adoption of improved management practices and are highly cost effective Fabregas et al., 2019; Ferdinand et al., 2021; Fabregas et al., 2024). We contribute to this literature by studying the delivery of short-run weather forecast information to farmers in at-scale experiment with a saturation design that permits identifying spillovers.

We build on prior work on saturation cluster designs and measurement under interference (Hudgens and Halloran, 2008; Baird et al., 2018 Vazquez-Bare, 2023). In applied settings, we contribute to the literature on information experiments at-scale that consider spillovers in a variety of settings including tax compliance (Cruces et al., 2024, political advertisements (Gerber et al., 2011; Enríquez et al., 2024, labor (Duflo and Saez, 2003; Crépon et al., 2013). Viviano and Rudder (2024) provide theoretical grounding for a saturation design with relatively few clusters, with the objective of identifying optimal saturation by detecting when spillovers level off. The method is to first assign clusters randomly to saturation levels (the standard procedure) and then randomly adjust the saturation level by a small perturbation (e.g., 5 or 10 %), preserving the average assignment in cluster groups while gaining the ability to estimate the marginal treatment effect of increasing or decreasing the treatment share across clusters, without decreasing power to detect main effects. Despite the improvement in welfare as the share treated increases, we use novel methods to show that spillovers level off after when between 50-70% of farmers are treated in a cluster. This means that policymakers under a budget constraint would have a higher impact by reaching more clusters without treating all individuals within the cluster rather than attempt to reach 100% of individuals within fewer clusters.

2 Background and Setting

Cotton is a politically important, high-value cash crop that contributes to Pakistan's domestic textile manufacturing. All of cotton production is sold through local intermediaries into Pakistan's $17 billion textile industry which accounts for 60% of export values.[2] Cotton growing is concentrated in the Indus River valley in Punjab and Sindh provinces. The majority of production (66%) is in Punjab province, concentrated in 11 out of 23 districts (USDA, 2023).[3] Cotton production is input intensive and grown under irrigation. Farmers use both government-supported irrigation canals and privately-operated tubewell systems. High quality weather forecasts enable farmers to plan important farming tasks - planting, irrigation dates, fertilizer, pesticides, herbicides, and fungicides.

In Punjab province, 1.6 million farmers grow cotton and the majority have farms that are less than 5 acres.[4] Because it is difficult to reach large numbers of farmers through traditional in-person agricultural extension, our NGO collaborator Precision Development (PxD) launched an agronomic advisory service for cotton growers in 2021. The program included a roster of 400,000 farmer phone-numbers located in 40 cotton-growing tehsils in Punjab Province, representing nearly one-third of all cotton growers in the province.[5] The roster of farmer phone numbers came from the Punjab Department of Agriculture and consisted of cotton farmers who signed up for government agricultural services. For the 2022 cotton growing season, the NGO expanded operations by adding a weather forecasting service. Both agronomic advisory and weather forecasts were delivered to farmers via automated voice calls. The phone number that appeared on calls was known by farmers to indicate message delivery from the Punjab Department of Agriculture. Farmers could opt out of messages by blocking the phone number or unsubscribing, although in practice this was not common.

2.1 Weather Forecasts and Information Sharing

General weather information is available to farmers with access to internet and media. Weather information is shared by radio, television, through smartphone apps and through the Pakistan Meterological Department. However, prior to the roll-out of weather forecasts phone calls, there was no systematic delivery of short-range weather forecasts by the government directly to farmers.

[2]Source: World Bank's World Integrated Trade Solution database. https://wits.worldbank.org/

[3]USDA Foreign Agriculture Service. https://ipad.fas.usda.gov/highlights/2023/08/Pakistan/index.pdf

[4]Agricultural Census by the Pakistan Bureau of Statistics.

[5]Tehsils are the second smallest administrative unit in Pakistan, equivalent to a county in the United States.

Pilot work with farmers indicated that there was a high-demand for short-range weather information and that farmers share weather information with each other. During pilot surveys before the full project launch, about half of farmers reported having access to weather information and 87% of farmers said they share it with other people in their village. Acting on information from peers and government sources is also relatively common. Among farmers who reported adopting any new practice in the prior year (25% of the sample), 30% reported learning general agronomic information from peers, compared to 47% learning from government extension officers, and 26% learning from agrodealers. Even without direct communication, farmers also learn from each other by observing behavior on neighboring plots, although we do not directly measure this behavior in our setting.

3 Research Design

We set up a cluster randomized experiment where we first allocate tehsils to a saturation group and then assign individual farmer phone numbers to treatment. Out of 40 eligible tehsils, the research team worked with a sub-sample of 35 tehsils because the NGO set-aside the remaining tehsils for a separate pilot project. The sample frame consists of nearly 400,000 existing users of a digital cotton farming advisory service. The eligible sample was allocated to treatment and control using a saturation design that varied the share of farmers per tehsil that were allocated to treatment. The control group received calls for standard cotton agronomic advisory. The treatment group received advisory calls plus 2-day weather forecast throughout the cotton growing season. We treat the tehsil as the cluster.[6]

Viviano and Rudder (2024) provide an approach to measure spillovers by varying treatment probability in the presence of unknown interference between units within clusters (e.g. farmers within tehsils). In many experiments, treatment is assigned by allocating a predetermined share of units within clusters and assuming that spillovers between treated and control units are minimal (SUTVA, or stable unit treatment value assumption). When researchers suspect that treatment effects may spillover to other units, they may randomly vary the share of units assigned within clusters to learn about the presence of information sharing via social learning (Miguel and Kremer, 2004; Hudgens and Halloran, 2008; Baird et al., 2018). We first randomly assigned pairs of clusters to a high, medium, and low probability of treatment. The high and medium staturation induce

[6]We use tehsils for clusters because tehsils were the narrowest reliable administrative unit reported in the government rosters. We also view tehsils as a reasonable unit to measure spillovers because they are small enough that at least some farms will likely know each other and interact but they are not too small such that information spillovers are too high, which would lead to high intra-cluster correlation and decrease power.

perturbation by 10% (positive and negative respectively) around 50% and 70% of individuals treated in two consecutive experimentation wave. Within each of these groups we also perturbated the assignment within pairs slightly (by 5%) to identify the improvement from a small change in the assignment probability. Viviano and Rudder provides econometric theory and proofs to show that this method is well-powered even when there are relatively few clusters, as in this setting with 35 tehsils. The assumption is that there are minimal differences based on observables and that information does not spread too fast between treated and untreated units. We verify balance on observables using administrative data shown by the table 1 in the appendix. Our set-up is a good setting for this assumption because we narrowly focus on cotton-growers, who have relatively similar land size, land quality, and access to irrigation.

By varying treatment intensity across a relatively large geography with many participants, our design will allow us to detect generalized social learning from an at-scale experiment of digital information delivery. We also leverage detailed information on phone call listening rates to weigh benefits of targeted delivery compared to mass delivery.

3.1 Assignment flow chart

There are two levels of random assignment: cluster and individual. Figure 1 shows the randomization diagram. In the first wave assignment, we varied treatment intensity at the tehsil level. In the second wave, we increased treatment intensity in all tehsils and increased the percentage of users assigned to treatment based on their uptake. For adaptive assignment, we vary the share of farmers assigned based on their historic uptake of the service. Users with higher uptake were 10 percentage points more likely to be assigned to treatment. We do not report results from the adaptive assignment in the main results. First, the second wave assignment took place later in the season than originally planned, therefore, those additionally treated phone numbers had less time with the service. Second, the likelihood of being assigned to treatment was only slightly higher for those with above median listening rates, and we were underpowered to detect effects. For our collaborator, it was important to provide weather forecasts to more people, but those rates were scaled within each grouping of low-medium-high saturation rates. We therefore focus on the continuous version of the saturation rate and the split between low-high saturation rates, as explained in the empirical estimation section. Saturation rates (and therefore propensity scores of assignment) vary over time from wave 1 to wave 2, but the low-medium-high split was randomized at the beginning of the season.

3.2 Data and Sample

We leverage three data sources in our analysis: administrative data the records whether participants listen to phone-calls, a phone survey throughout the 2022 cotton-growing season, and an endline survey with a subset of participants 8 months after the end of the project in July 2023.

3.2.1 Administrative data

Administrative data provided by the NGO provide information on all 460,000 phone numbers provided by the government roster. It includes records for dates when messages were delivered, the length of the phone message, and the length of the phone message listened to by the participant for both weather forecasts and agronomic advisory. Administrative data are available for both the pre-treatment season which took place from May to October, 2021, and during the project season which took place from May to October, 2022.

3.2.2 2022 Mid-season phone survey

During the 2022 agricultural season, we implemented a phone survey with a sample of farmers drawn from the administrative data records. The primary goal of the survey was to collect real-time data on weather responses from different farmers throughout the season. The primary outcomes collected during the phone survey was to elicit weather predictions for the next day, recall information for weather in the prior day, the exact days of recent farming activities (irrigation, fertilizer, pesticide applications), information about sharing weather, and expectations for yields during harvest. The survey firm had a capacity to call up to 400 farmers per week to complete a survey. We decided to draw a repeat cross section of new farmers each week instead of a repeat panel of the same farmers because we wanted to maximize geographic coverage and minimize survey bias effects that could arise from repeat surveys to the same set of farmers. At the end of the season, we have surveys for 5,700 farmers, of which 3,600 were cotton-growers. For most analysis we use the full sample of farmers, but we only asked details on agricultural inputs for cotton-growers.

For weekly phone surveys, We drew random lists of farmer phone numbers to be approximately 50% treatment and 50% control in order to optimize power. We also stratified call lists based on tehsil locations to ensure that approximately equal numbers of users were called from each tehsil for each weekly survey. Over the course of the season, the survey company attempted to call 17,600 users, of which 6,500 completed a survey (5,700 for our sample), resulting in a survey completion rate of 37%. Although this is low, it is in line with phone survey response rates in some other low- and middle-income settings (Gibson et al., 2019, Henderson et al., 2020; Gourlay et al., 2021). In Appendix 3, we show how the survey population differs from the full roster of phone numbers.

3.2.3 2023 Endline phone survey

Approximately eight months after the mid-season phone survey, we collected another phone survey with 2,000 respondents. These respondents were a randomly drawn subsample from the original phone survey, balanced on treatment status and location. We asked a series of follow-up questions related to cotton productivity, including yields, profits, costs, and timing of harvest. We also asked farmers about their planting behavior in the subsequent season.

3.3 Main Outcomes

We pre-registered the following outcomes: engagement with weather and advisory phone calls, weather forecast predictions, sharing of information, timing of input applications, yields, and input use and profits. In addition, we pre-registered short and long-run expectations and beliefs as secondary outcomes and pre-specified forecast accuracy and weather trends (shocks and extreme weather events) as a source of heterogeneity. Before the second survey, we registered short-run adaptation (crop, seed choices, conservation agriculture practices) and perceptions of climate change as additional outcomes.

Forecast predictions and recall: Messages of forecasts sent to mobile phones included an estimate of the expected amount of rain in millimeters, the maximum temperature, and the minimum temperature in celsius. Forecasts did not include probabilistic information. Instead the forecast provider worked with the NGO to establish thresholds that would define whether forecasts would positively report rain levels. During the mid-season phone survey, farmers their predictions for weather for the day following the survey. For rainfall, farmers were asked "Do you think it will rain tomorrow?" For temperature, farmers were asked, "What do you think the max(min) temperature will be tomorrow?" We then compared farmers' predictions with the forecast information provided on those days. For rainfall, we define a binary variable equal to one if the farmer's rain prediction aligns with the forecast (e.g. if a farmer predicts any rain level above zero and the forecast also reports any rain level above zero, they are assigned one; zero otherwise). For temperature, we calculated the absolute deviation of predicted temperature minus the forecast temperature so that predictions above and below the forecast are treated similarly. Negative values in the temperature prediction indicate that the farmer's prediction was closer to the forecast.

Sharing information: The first variable in this outcome grouping is a binary variable equal to one if farmers reported learning weather information from any official source over the prior week. Second, farmers were asked if they shared weather information with any peers. Finally, farmers were asked if any peers shared weather information with them.

Farming behavior: During each mid-season phone survey, farmers were asked if they carried out specific farming tasks and the precise date of those tasks. The primary tasks are planting, irrigation, fertilizing, adding pesticides, herbicides, and fungicides. Not all farmers completed all tasks during a given week, so sample sizes vary based on whether a task was reported during the phone survey. We then compared the dates of tasks with both the realized and forecast weather for the same day. For the main results, we report results for realized weather, and report results for forecasts in the appendix.

3.4 Balance

We report two balance tables: individual balance of the phone survey sample, cluster balance of phone survey participants. Table 1 reports means, standard deviations and differences in means tests for baseline covariates for the sample of farmers drawn into the mid-season phone survey. Table 2 reports covariate balance based on whether phone survey farmers were allocated to either low or high saturation tehsils. We see a difference in household size, where high saturation households are slightly smaller, but are otherwise similar based on other characteristics (education, age, gender, education, prior season yields, farm size) and participation (similar listening rate to advisory services in the prior year). We also see that there are differences based on measures of minimum temperature stress. This is likely driven by a few outliers because temperature stress based on minimum temperature (defined as below 15 degrees Celsius, the lower bound for cotton) was not common. Tehsil saturation is balanced on forecast accuracy for rain and maximum temperature, which is used for heterogeneity analysis.

In addition, Table 3 reports sample differences for the group of phone numbers drawn into the phone survey sample compared to the full sample of phone numbers provided by the government. Phone survey participants are largely similar to the administrative population, except that they are more likely to listen to advisory service messages.

4 Estimation

4.1 Intention to Treat and Total Spillover

To estimate intention to treat and total spillover effects, we use the following specification:

$$Y_{ic} = \alpha + \beta_1 Treat_{ic} + \beta_2 Saturation_c + \epsilon_{ic} \tag{1}$$

Y_{ic} denotes the outcome for farmer i in cluster c. Standard errors are clustered at the tehsil-level. Recall that we had a two stage randomization procedure where clusters were first randomly assigned to a saturation level and then individuals within each cluster were randomly assigned to a treatment status. β_1 is the intent-to-treat effect for individuals assigned to treatment. β_2 captures the total spillover effect, and is composed of spillovers for treated and untreated units. In other words, it includes both the additional effect of being treated along with many treated peers, and the spillover effect of untreated units that have many treated peers (Hudgens and Halloran 2008; Baird et al., 2018). In this specification, in wave 1 treatment saturation is specified as a continuous variable that ranges from probabilities $p_L = \{0.05, 0.10, 0.15\}$ for low saturation, $p_M = \{0.35, 0.45\}$ for medium saturation, and $p_H = \{0.55, 0.65\}$ for high saturation. In the second wave, treatment saturation was assigned as $p_L = \{0.10, 0.15, 0.20\}$ for low saturation, $p_M = \{0.55, 0.65\}$ for medium saturation, and $p_H = \{0.75, 0.85\}$ for high saturation.[7]

4.2 Spillover Effects on Treated and Untreated Units

To disentangle the spillover effects on treated and untreated units, we use the following saturated regression:

$$\begin{aligned} Y_{ic} = \alpha &+ \beta_1 Treat \times LowSaturation_{ic} + \beta_2 Control \times HighSaturation_c \\ &+ \beta_3 Treat \times HighSaturation_c + \epsilon_{ic} \end{aligned} \tag{2}$$

For ease of interpretation, we combine the medium and high saturation groups into one group (called High) to compare against the Low Saturation group. The reference category is control units

[7] The two wave assignment strategy generates differences in treatment duration for individuals in the sample. In our main analysis, we do differentiate between the two assignment waves. In practice, wave 2 was implemented relatively late in the season. Results are similar when dropping those who received treatment later in the season.

in the low saturation clusters (we have no pure control clusters). β_1 identifies the effect on treated units in low saturation clusters, β_2 identifies the spillover effect on untreated control units located in high saturation clusters, and β_3 identifies the treatment effect of directly treated units in high saturation clusters. The difference between β_1 and β_3 represents the spillover on treated units.

5 Results

5.1 Weather Predictions

Columns 1, 3, and 5 in Table 4 reports treatment effects regressions following specifications in equations 1 and columns 2, 4, and 6 report results using equation 2. There are three measures based on whether the farmer prediction aligns with forecasts - rainfall, maximum temperature, and minimum temperature, as described in Section 3.3. Forecast recipients who are directly treated have rainfall predictions that are 3.4 percentage points more likely to be aligned with forecasts. The total spillover effect exhibits a positive and significant relationship, a 1% increase in the tehsil share treated increases the probability that farmer predictions align with forecasts by 10.2 percentage points. Results are similar for predictions of maximum and minimum temperature - directly treated units have more aligned predictions and the spillover effect is larger in magnitude. Results from the saturated regressions in columns 2, 4, and 6 suggest that the spillover on treated units has the largest effect, compared to the treated units in low saturation clusters and control units in high saturation clusters.

Non-response for some weather forecast variables was high for maximum and minimum temperature. Only about one quarter of phone survey participants supplied a response in degrees Celsius that could be converted into a measure of absolute deviation. We think this is because without access to a granular forecasts, people have fewer tools to predict temperature with precision. For rainfall, the question was simpler to answer. Participants were asked a yes or no question, "do you think it will rain tomorrow?" which many more respondents were willing to answer. Table 10 reports results on the likelihood of answering these questions. We see that treated farmers were more likely to provide answers to these questions (columns 1 and 5). As an alternative specification, we impute group-level means and re-run regressions with the full sample (columns 3 and 7) to increase precision in estimates from the underrepresented control group and see that results are similar. While the differential non-response is a concern, we highlight these results to show evidence that treated farmers attended to the information and updated their beliefs.

5.2 Receiving and Sharing Weather Information

Table 5 reports outcomes related to receiving and sharing weather information. In this setting, it is possible that control respondents have access to weather information from other sources besides the forecasts provided by the collaborator. Column 1 reports the probability that a respond reports learning a weather forecast from an official source in the previous week. About 51% of the control group received forecasts, and treated respondents were 12 percentage points more likely to report receiving forecasts. The most common other sources were internet/smartphone, TV, and radio. The intervention induced treated farmers to share weather information with their peers (columns 3 and 4). About 40% of control farmer share weather information and treated farmers 10-13 percentage points more likely to share - a 25% increase. Spillover effects in column 6 are particularly relevant - control farmers in high saturation tehsils were 14% more likely to report receiving weather information from peers, despite not having learned information from official sources (column 2). This validates information sharing as a plausible channel for the spillover effects observed in farming behavior described in the subsequent section.

5.3 Short-run Farming Behavior

We focus on two types of farming responses: 1. avoiding rain, and 2. avoiding heat for a series of farming tasks - planting, irrigation, fertilizing, applying pesticides, fungicides, and herbicides, and an index of all 6 tasks. Panel A of Table 7 reports intent-to-treat and total spillovers and Panel B reports spillovers as a saturated regression for avoiding rain using realized weather. Table 7 is organized similarly for avoiding heat using realized weather. In the appendix, Tables 11 and 12 report results using forecasts to establish avoidance behavior and results are similar. Table 6 shows that treated farmers are more likely to avoid rain when irrigating, fertilizing, and applying pesticides, and that there are large spillovers, particularly for irrigation and fertilizing for both control and treated units in high saturation tehsils (columns 2 and 3). Focusing on Panel B, twenty-two percent of control farmers in low saturation tehsils use irrigation on days without rain, treated farmers in low-saturation tehsils are 14% more likely to irrigate on days without rain, and control and treated units in high saturation tehsils are respectively 47% and 68% more likely irrigate on days without rain (point estimates in percentage points are .032, .104, and .151, respectively). Most farmers in this sample use a combination of public and private sources of irrigation, so that this type of reallocation is feasible without renegotiating irrigation time with neighbors on a shared system.

For fertilizing, control farmers in high saturation areas are 30% more likely, and treated respondents in high saturation are 40% more likely to avoid rain.[8] Recall that in Table 2, we show that tehsils in low and high saturation are balanced in terms of seasonal rainfall patterns and rainfall accuracy, which reduces concerns that highly saturated tehsils systematically differed in terms of rainfall patterns. We see similar patterns for pesticide applications and for the index in column 7. Fungicide and herbicide exhibit noisier and less clear overall patterns, but fewer farmers applied these inputs overall. It appears that treated farmers were less likely to avoid rain when planting, but this was early in the season after farmers only had been receiving forecasts for a few weeks. If farmers planted before treated began, they are excluded, which is why the sample size for planting is lower than for irrigation or fertilizing.

Table 7 reports analogous results for avoiding heat, defined as days with maximum temperature above 37C. Results are statistically weaker, but exhibit similar patterns across all tasks: treated farmers in low and high saturation tehsils as well as control units in high saturation tehsils all avoid heat to a greater extent than control units in low saturation tehsils. One major difference is that control units in low saturation tehsils are 'better' at avoiding heat compared to avoiding rain: the share of control farmers that avoid heat ranges from 81-91% for most tasks, compared to 20-30% for avoiding rain. Nonetheless, treated farmers in high saturation areas are 9% more likely to avoid heat when irrigating, 8% more likely during fertilizing, and 12% more likely to avoid heat when applying fungicides.

5.4 Productivity Outcomes

Panel A of Table 8 reports spillover effects for productivity outcomes, including yield, profits, revenue, input costs, output prices, and harvest timing. Column 1 shows that yields for treated farmers are 5% higher for all treated farmers and 3-7% larger in saturated regressions. Results are only marginally significant for the treated farmers in low saturation tehsils (Panel B, column 1). In both panels, most measures are too noisy to reject differences from zero, but patterns suggest that profit decreases as saturation rates increase and revenue slightly increases for treated farmers, but not spillover farmers. In both Panel A and B, input costs rise as saturation rates increase, and is statistically larger for treated farmers in high saturation tehsils whose expenditure increased by

[8] Some types of common fertilizer are water soluble and benefit from some rain to dissolve. We lack precise information about the types of fertilizers used to inspect this in detail but argue that farmers would still prefer to apply fertilizers on days without rain and use irrigation to dissolve fertilizers as needed.

12%.

Columns 5 and 6 report the output prices for cotton and the harvest timing. Although measured with noise, we see that prices are 1.6-3.8% lower in higher saturation tehsils, suggesting that one reason why profits are lower is that farmers received lower farmgate prices when selling cotton. Treated farmers in low saturation tehsils harvest their crop later by a few weeks - control farmers harvest in late September and treated farmers in low saturation tehsils harvest in early October. These patterns suggest that there may high saturation may lead to some crowding out at the time of harvest. Harvest is labor-intensive and 90% of farmers hire labor during the period. Treated farmers in low saturation tehsils may have had an information advantage to delay harvests which afforded slightly better prices and overall better profits.

In Panel B, input costs rose by 9-13%, which was statistically different than zero. Columns 5 and 6 show suggestive evidence that treated farmers harvested later and received lower prices, which is a possible mechanism to explain lower overall profits. Panel B reports input costs per acre and shows that costs increased by 9-16% for irrigation, pesticides, and fertilizer. Given that farmers responded to treatment by adjusting the timing of their input applications, it suggests that some of these adjustments were costly in the short-run.

5.5 Forecast Accuracy and Take-up

Among participants in the 2022 phone survey, 75% listened to at least one weather message and 77% listened to at least one advisory message. On average, people who listened to at least one weather message listened to 30% of all messages sent, highlighting that people selectively answered messages throughout the season. Table 9 reports differences in message take-up rates by different measures of forecast accuracy. First, satellite rainfall accuracy (row 1) refers to the share of days where rainfall forecasts aligned with realized rainfall according to satellite measures. Farmer-reported rainfall accuracy (2) is the share of days when whether forecasts aligned with farmer-reported rainfall during the survey. Absolute rainfall deviations (3) is the average absolute deviation of forecasts from realized weather, measured in millimeters. Rainfall false alarms (4) is the number of days with false alarms in precipitation, defined as days where forecasts predicted no rain and a heavy rain fell, or when forecasts predicted heavy rain and no rain fell. Rainfall RMSE (5), max temp (6), and min temp (7) are the root mean-squared error for rainfall, max and min temp from a tehsil-level regression of the realized value on the forecast value. Measures 3-7 are common measures used in meteorology to assess forecast accuracy.

Measures 1 and 2 were constructed as binary measures that can be interpreted as the share of days where a forecast of positive rain corresponded to realized rain, and vice versa for forecasts and realized weather without rain. Measure 1 uses satellite measures of realized rain compared to forecasts. Measure 2 uses farmer-reported rainfall. During the phone survey, respondents were asked if it rained the prior day and we matched that up with forecast on the same day. Measure 2 is the 'ground truth' data that best captures whether forecasts were accurate for participants. According to both measures, increased accuracy is correlated with higher average take-up, suggesting that farmers in areas with lower accuracy decreased their take-up as a response. A 10% increase in the share of days with an accurate forecast leads to a 1.6-2.9% increase in average take-up. A few of the other satellite-based measures of accuracy show the opposite trend; an increase in rainfall deviations between forecast and realized weather (row 3) and an increase in max temp RMSE (row 6) are related to a an increase in take-up (by construction, higher values of measures 3-7 imply *lower* accuracy). However, these results are not necessarily inconsistent, measures 1 and 2 are binary measures while 3 and 6 are continuous. It is possible that higher rainfall deviations correspond to days with rain, and the binary measure of 'yes' or 'no' rain are what matters. For temperature, overall RMSE was low across tehsils and in general temperature forecasts tend have higher accuracy because they are more spatially correlated compared to rainfall.

6 Conclusion

After implementing a saturation design at large-scale with 400,000 farmers in 35 tehsils, we show that information spillovers are common and influence farmer decision making. We separately identify intent-to-treat effects, total spillovers, as well as spillovers on treated and untreated farmers. We show a pattern of behavior change that indicates that spillovers occur for both treated and control farmers, indicating that treated farmers benefit from from having other treated farmers to discuss farming and that control farmers benefit indirectly through information sharing.

References

Ahmad, H. F., Gibson, M., Nadeem, F., Nasim, S., and Rezaee, A. (2022). Forecasts: Consumption, production, and behavioral responses.

Baird, S., Bohren, J. A., McIntosh, C., and Özler, B. (2018). Optimal design of experiments in the presence of interference. Review of Economics and Statistics, 100(5):844–860.

Bandiera, O. and Rasul, I. (2006). Social networks and technology adoption in northern mozambique. The economic journal, 116(514):869–902.

Barnett-Howell, Z. (2021). No small potatoes: Agricultural risk and investment under uncertainty. Working paper.

Burlig, F., Jina, A., Kelley, E. M., Lane, G. V., and Sahai, H. (2024). Long-range forecasts as climate adaptation: Experimental evidence from developing-country agriculture. Working paper, National Bureau of Economic Research.

Camacho, A. and Conover, E. (2019). The impact of receiving sms price and weather information on small scale farmers in colombia. World development, 123:104596.

Cole, S., Harigaya, T., and Surendra, V. (2023). Probabilistic weather forecasts and farmer decision making in rural india.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment. The quarterly journal of economics, 128(2):531–580.

Cruces, G. A., Tortarolo, D., and Vazquez-Bare, G. (2024). Design of partial population experiments with an application to spillovers in tax compliance. Documentos de Trabajo del CEDLAS.

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly journal of economics, 118(3):815–842.

Enríquez, J. R., Larreguy, H., Marshall, J., and Simpser, A. (2024). Mass political information on social media: Facebook ads, electorate saturation, and electoral accountability in mexico. Journal of the European Economic Association, page jvae011.

Fabregas, R., Kremer, M., Lowes, M., On, R., and Zane, G. (2024). Digital information provision and behavior change: Lessons from six experiments in east africa. Working paper, National Bureau of Economic Research.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. Science, 366(6471):eaay3038.

Fafchamps, M. and Minten, B. (2012). Impact of sms-based agricultural information on indian farmers. The World Bank Economic Review, 26(3):383–414.

Ferdinand, T., Illick-Frank, E., Postema, L., Stephenson, J., Rose, A., Petrovic, D., Migisha, C., Fara, K., Zebiak, S., Siantonas, T., et al. (2021). A blueprint for digital climate-informed advisory services: Building the resilience of 300 million small-scale producers by 2030. Working paper.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. Journal of political Economy, 103(6):1176–1209.

Fosu, M., Karlan, D., Kolavalli, S., and Udry, C. (2018). Disseminating innovative resources and technologies to smallholder farmers in ghana: Results of the weather forecasting program. Technical report, Innovations for Poverty Action Preliminary Results Brief.

Gerber, A. S., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? results from a randomized field experiment. American Political Science Review, 105(1):135–150.

Gibson, D. G., Wosu, A. C., Pariyo, G. W., Ahmed, S., Ali, J., Labrique, A. B., Khan, I. A., Rutebemberwa, E., Flora, M. S., and Hyder, A. A. (2019). Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in bangladesh and uganda. BMJ global health, 4(5):e001604.

Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P., and Zezza, A. (2021). High-frequency phone surveys on covid-19: good practices, open questions. Food policy, 105:102153.

Henderson, S., Rosenbaum, M., Roemer, J., Glazerman, S., Parkerson, D., and Warren, S. (2020). Remote surveying in a pandemic: research synthesis. Innovation for Poverty Action.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. Journal of the American Statistical Association, 103(482):832–842.

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. Nature, 529(7584):84–87.

Linsenmeier, M. and Shrader, J. G. (2023). Global inequalities in weather forecasts. Working paper.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1):159–217.

Molina, R. and Rudik, I. (2023). The social value of predicting hurricanes. Working paper, National Bureau of Economic Research.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the indian green revolution. Journal of development Economics, 73(1):185–213.

Rosenzweig, M. R. and Udry, C. R. (2019). Assessing the benefits of long-run weather forecasting for the rural poor: Farmer investments and worker migration in a dynamic equilibrium model. Working paper, National Bureau of Economic Research.

Shrader, J. G., Bakkensen, L., and Lemoine, D. (2023). Fatal errors: The mortality value of accurate weather forecasts. Working paper, National Bureau of Economic Research.

Vazquez-Bare, G. (2023). Identification and estimation of spillover effects in randomized experiments. Journal of Econometrics, 237(1):105237.

Viviano, D. and Rudder, J. (2024). Experimental design under network interference. arXiv preprint arXiv:2003.08421.

Wollburg, P., Markhof, Y., Bentze, T., and Ponzini, G. (2024). Substantial impacts of climate shocks in african smallholder agriculture. Nature Sustainability, pages 1–10.

Yegbemey, R. N., Bensch, G., and Vance, C. (2023). Weather information and agricultural outcomes: Evidence from a pilot field experiment in benin. World Development, 167:106178.

7 Tables and Figures

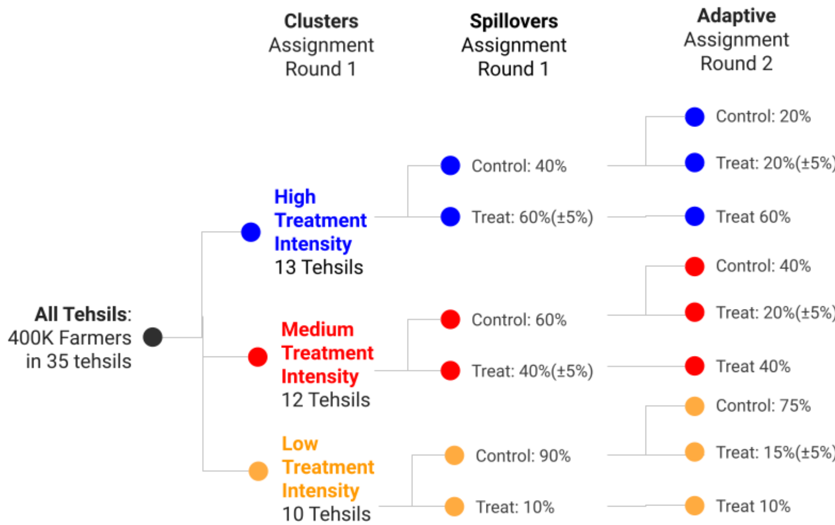

Figure 1: Saturation Design

Experimental Balance

Table 1: Experimental Balance by Individual Treatment Status

Variable	(1) Control Mean/SD	(2) Treatment Mean/SD	T-test Difference Control-Treat
Woman	0.042	0.031	0.011***
	(0.201)	(0.186)	
Age	42.100	41.914	0.186
	(16.420)	(14.841)	
Education	5.970	5.876	0.094
	(7.086)	(8.189)	
Household size	5.464	5.413	0.051
	(4.022)	(3.284)	
Owns land	0.868	0.877	-0.010
	(0.654)	(0.613)	
Farm size (acres)	5.578	5.522	0.055
	(5.925)	(6.454)	
Uses smartphone (self-reported)	0.342	0.354	-0.012
	(0.616)	(0.434)	
Uses WhatsApp (automated)	0.322	0.331	-0.009
	(0.464)	(0.619)	
Prior season yields (mounds/acre)	25.938	25.666	0.272
	(13.403)	(15.319)	
Prior season advisory listening rate on all calls	0.566	0.580	-0.015
	(0.675)	(0.725)	
Prior season advisory listen at least once	0.936	0.942	-0.006
	(0.286)	(0.291)	
Completed Farming Task Module	0.611	0.604	0.007
	(1.023)	(1.085)	
N	2547	3187	
Clusters	35	35	
F-test of joint significance (F-stat)			3.479***
F-test, number of observations			5734

Table 2: Experimental Balance by Cluster Saturation

Variable	(1) Low Saturation Mean/SD	(2) High Saturation Mean/SD	T-test Difference Low-High
Woman	0.036 (0.253)	0.036 (0.208)	-0.000
Age	41.512 (16.997)	42.046 (15.271)	-0.534
Education	5.735 (8.530)	5.949 (9.692)	-0.214
Household size	6.136 (6.791)	5.294 (3.966)	0.842***
Owns land	0.883 (0.474)	0.871 (0.797)	0.012
Farm size (acres)	5.207 (9.518)	5.611 (7.915)	-0.404
Uses smartphone (self-reported)	0.335 (0.318)	0.350 (0.512)	-0.016
Uses WhatsApp (automated)	0.312 (0.509)	0.329 (0.588)	-0.017
Prior season yields (mounds/acre)	25.683 (15.714)	25.803 (16.484)	-0.120
Prior season advisory listening rate on all calls	0.558 (0.688)	0.576 (0.841)	-0.018
Prior season advisory listen at least once	0.942 (0.270)	0.939 (0.330)	0.003
Completed Farming Task Module	0.587 (1.427)	0.609 (1.313)	-0.022
Days with max temp stress	67.421 (423.113)	49.711 (548.746)	17.710
Days with min temp stress	0.148 (6.337)	16.357 (426.056)	-16.209**
Days with max temp extreme	7.710 (139.385)	5.793 (137.749)	1.917
Days with min temp extreme	0.000 (0.000)	4.430 (205.925)	-4.430
Cumulative Rainfall	686.679 (682.429)	691.990 (1151.743)	-5.311
Days with extreme rain	18.938 (27.176)	18.701 (41.437)	0.237
Days with any rain	110.879 (130.830)	108.540 (219.532)	2.339
Days with heavy rain	28.896 (49.395)	27.355 (54.634)	1.541
Days with no rain	72.121 (130.830)	74.460 (219.532)	-2.339
Forecast accuracy - rain	8.528 (10.688)	9.118 (19.156)	-0.590
Forecast accuracy - max temp	1.742 (5.314)	1.870 (3.421)	-0.128
Forecast accuracy - min temp	1.130 (2.355)	1.447 (5.665)	-0.317***
N	1608	4126	
Clusters	10	25	
F-test of joint significance (F-stat)			0.696
F-test, number of observations			5734

Table 3: Difference in Means for Administrative Data and Phone Survey Sample

	Administrative Data Mean (sd)	Completed Phone Survey Mean (sd)	Test
Woman	0.043 (0.204)	0.034 (0.182)	0.015
Age	51.132 (14.455)	50.918 (13.644)	0.419
Education	5.457 (4.309)	5.949 (4.237)	<0.001
Household size	5.404 (3.303)	5.506 (3.177)	0.092
Owns land	0.852 (0.355)	0.872 (0.334)	0.002
Farm size (acres)	5.193 (4.480)	5.508 (4.718)	<0.001
Uses WhatsApp (automated)	0.267 (0.443)	0.323 (0.467)	<0.001
Prior season advisory listening rate on all calls	0.426 (0.346)	0.575 (0.317)	<0.001
Prior season advisory listen at least once	0.786 (0.410)	0.939 (0.239)	<0.001
Weather listening rate	0.125 (0.223)	0.185 (0.262)	<0.001
Pick up weather at least once	0.438 (0.496)	0.512 (0.500)	<0.001
Advisory listening rate	0.208 (0.253)	0.327 (0.274)	<0.001
Pick up advisory at least once	0.774 (0.418)	0.958 (0.200)	<0.001
Treat	0.611 (0.488)	0.556 (0.497)	<0.001
N	442,783 (98.7%)	5,734 (1.3%)	

Main Results Tables

Table 4: ITT and Spillover Effects of Weather Predictions

	Rainfall		Max Temperature		Min Temperature	
	(1) Prediction equals forecast (0/1)	(2) Prediction equals forecast (0/1)	(3) Absolute Deviation from Forecast	(4) Absolute Deviation from Forecast	(5) Absolute Deviation from Forecast	(6) Absolute Deviation from Forecast
Treat	0.034***		-0.676***		-1.074***	
	(0.012)		(0.172)		(0.232)	
Tehsil Share Treated	0.102***		-1.026**		-1.033**	
	(0.035)		(0.426)		(0.412)	
Treat×Low Saturation		0.023		-0.486		-1.499***
		(0.023)		(0.475)		(0.493)
Control×High Saturation		0.008		-0.224		-0.698*
		(0.027)		(0.489)		(0.381)
Treat×High Saturation		0.051*		-1.003**		-1.720***
		(0.025)		(0.471)		(0.310)
Pvalue: Treat(Low)=Control(High)		0.5188		0.3728		0.0839
Pvalue: Treat(Low)=Treat(High)		0.2682		0.0850		0.6005
Pvalue: Control(High)=Treat(High)		0.0086		0.0002		0.0004
Control Mean	0.438	0.432	4.056	4.227	4.915	5.446
Observations	5144	5144	1283	1283	1189	1189
Adj R-Squared	0.004	0.001	0.019	0.015	0.026	0.024

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent's predicted forecast for rainfall, maximum temperature, and minimum temperature for the following day align with the forecasts provided by phone messages. Sample sizes differ due to non-response. Results are similar if group means are imputed as reported in Table 10 in the appendix.

Table 5: ITT and Spillover Effects of Receiving and Sharing Weather Information

	(1) Received Weather Info from Official Sources	(2) Received Weather Info from Official Sources	(3) Shared Weather Info w/ Peers	(4) Shared Weather Info w/ Peers	(5) Peers Shared Weather Info	(6) Peers Shared Weather Info
Treat	0.119***		0.105***		0.029*	
	(0.014)		(0.012)		(0.015)	
Tehsil Share Treated	-0.035		-0.006		-0.015	
	(0.044)		(0.041)		(0.033)	
Treat×Low Saturation		0.108***		0.131***		0.113***
		(0.034)		(0.028)		(0.025)
Control×High Saturation		-0.010		0.015		0.048**
		(0.028)		(0.025)		(0.019)
Treat×High Saturation		0.109***		0.113***		0.056***
		(0.028)		(0.026)		(0.017)
Pval: T(Low)=C(High)		0.0034		0.0004		0.0111
Pval: T(Low)=T(High)		0.9710		0.5405		0.0082
Pval: C(High)=T(High)		0.0000		0.0000		0.5738
Control Mean	0.509	0.516	0.399	0.387	0.395	0.358
Observations	5733	5733	5732	5732	5723	5723
Adj R-Squared	0.014	0.013	0.011	0.011	0.001	0.002

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent received weather information from an official source such as government, TV, Radio, Mobile phone (Columns 1 and 2), whether they shared weather information with peers (Columns 3 and 4), and whether peers shared weather information with the respondent over the prior week (Columns 5 and 6). Columns 1, 3, and 5 report ITT treatment effects and total spillover effects as a continuous measure equal to the share of units treated in each tehsil. Columns 2, 4, and 6 report saturated regression where the reference category is control units in low saturation clusters. All outcomes are binary.

Table 6: ITT and Spillover Effects on Farming Behavior: Avoiding Rain

Panel A: ITT and Spillover Effects as Continuous Measure

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Rain Index
Treat	-0.036*	0.038**	0.010	0.029*	0.005	-0.011	0.003
	(0.020)	(0.015)	(0.015)	(0.016)	(0.027)	(0.021)	(0.017)
Tehsil Share Treated	-0.062	0.247***	0.127*	0.126*	-0.047	0.015	0.051
	(0.108)	(0.081)	(0.064)	(0.067)	(0.078)	(0.083)	(0.075)
Control Mean	0.552	0.256	0.190	0.235	0.259	0.264	0.000
Observations	2257	3303	3223	3046	1167	2071	3482
Adj R-Squared	0.001	0.019	0.005	0.006	-0.001	-0.001	0.000

Panel B: Spillover Effects as Saturated Regression

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Rain Index
Treat×Low Saturation	-0.015	0.032*	0.005	0.022	0.026	-0.024	0.005
	(0.020)	(0.017)	(0.018)	(0.021)	(0.036)	(0.023)	(0.017)
Control×High Saturation	0.067	0.104***	0.053**	0.051	-0.023	-0.011	0.045
	(0.064)	(0.034)	(0.023)	(0.031)	(0.038)	(0.047)	(0.038)
Treat×High Saturation	-0.064*	0.151***	0.070**	0.087**	-0.044	0.007	0.041
	(0.036)	(0.027)	(0.029)	(0.033)	(0.028)	(0.039)	(0.042)
Pval: T(Low)=C(High)	0.2339	0.0534	0.0932	0.3557	0.3129	0.7783	0.3160
Pval: T(Low)=T(High)	0.2382	0.0002	0.0467	0.0507	0.0949	0.4316	0.4083
Pval: C(High)=T(High)	0.0064	0.0728	0.4575	0.1239	0.5721	0.6963	0.9403
Control Mean	0.540	0.223	0.174	0.217	0.267	0.267	0.000
Observations	2257	3303	3223	3046	1167	2071	3482
Adj R-Squared	0.002	0.017	0.005	0.005	0.001	-0.001	0.000

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent carried out farming tasks on days without rain according to satellite measures of realized weather. Panel A reports ITT effects as a binary variable and spillover effects as a continuous measure set to the saturation level of each tehsil. Panel B reports saturated regression with the reference category of control units in low saturation tehsils. Panel B also reports t-tests of equality for each saturation by treatment pair. Sample sizes differ because responses are conditional on having reported a date for each task.

Table 7: ITT and Spillover Effects on Farming Behavior: Avoiding Heat

Panel A: ITT and Spillover Effects as Continuous Measure

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Heat Index
Treat	0.009	0.024**	0.016	0.020**	0.029	0.045***	0.036**
	(0.020)	(0.012)	(0.010)	(0.009)	(0.024)	(0.015)	(0.013)
Tehsil Share Treated	0.220	0.125	0.106	0.072	0.171*	0.165	0.125
	(0.289)	(0.090)	(0.082)	(0.082)	(0.094)	(0.114)	(0.135)
Control Mean	0.437	0.865	0.880	0.918	0.838	0.771	0.000
Observations	2293	3303	3225	3046	1170	2081	3482
Adj R-Squared	0.007	0.010	0.007	0.006	0.014	0.011	0.006

Panel B: Spillover Effects as Saturated Regression

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Heat Index
Treat×Low Saturation	-0.007	0.039**	0.015	0.041***	0.037	0.047**	0.040**
	(0.023)	(0.016)	(0.014)	(0.013)	(0.034)	(0.017)	(0.018)
Control×High Saturation	0.059	0.078*	0.053	0.028	0.075	0.056	0.054
	(0.166)	(0.042)	(0.040)	(0.038)	(0.051)	(0.070)	(0.090)
Treat×High Saturation	0.132	0.078*	0.068*	0.028	0.100**	0.098	0.091
	(0.175)	(0.040)	(0.036)	(0.038)	(0.049)	(0.064)	(0.085)
Pval: T(Low)=C(High)	0.6992	0.3494	0.3737	0.7036	0.5053	0.8978	0.8694
Pval: T(Low)=T(High)	0.4387	0.3179	0.1607	0.6990	0.2125	0.4264	0.5378
Pval: C(High)=T(High)	0.0752	0.9985	0.1746	0.9898	0.2290	0.1189	0.2876
Control Mean	0.427	0.841	0.865	0.908	0.813	0.757	0.000
Observations	2293	3303	3225	3046	1170	2081	3482
Adj R-Squared	0.006	0.010	0.007	0.003	0.010	0.006	0.004

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent carried out farming tasks on days without extreme heat above 37C according to satellite measures of realized weather. Panel A reports ITT effects as a binary variable and spillover effects as a continuous measure set to the saturation level of each tehsil. Panel B reports saturated regression with the reference category of control units in low saturation tehsils. Panel B also reports t-tests of equality for each saturation by treatment pair. Sample sizes differ because responses are conditional on having reported a date for each task.

Table 8: Treatment Effects on Farming Productivity

Panel A: ITT and Spillover Effects as Continuous Measure

	(1) Yields	(2) Profits	(3) Revenue	(4) Input Costs	(5) Output Price	(6) Harvest Month
Treat=1	0.603	1705.830	4279.998	2227.718	21.413	0.038
	(0.592)	(4926.933)	(5168.249)	(1871.689)	(117.490)	(0.061)
Tehsil Share Treated	-0.284	-15120.842	-7830.191	9887.454	-124.340	-0.021
	(2.540)	(16146.060)	(19157.336)	(5893.545)	(607.263)	(0.156)
Control Mean	11.780	25005.653	94968.054	69699.634	7703.776	9.829
Observations	1403	1386	1386	1408	1386	1323
Adj R-Squared	0.000	0.001	-0.000	0.005	-0.000	-0.002

Panel B: Spillover Effects as a Saturated Regression

	(1) Yields	(2) Profits	(3) Revenue	(4) Input Costs	(5) Output Price	(6) Harvest Month
Treat in Low	0.815*	921.211	5843.102	4967.618	12.674	0.251**
	(0.451)	(5255.108)	(5814.270)	(3219.009)	(265.888)	(0.119)
Control in High	0.317	-5932.430	-1496.390	5724.317	-298.533	0.118
	(1.456)	(8272.195)	(9523.803)	(3804.726)	(334.257)	(0.093)
Treat in High	0.796	-4455.690	2313.280	7626.130**	-190.940	0.085
	(1.420)	(7399.278)	(8798.241)	(3721.279)	(344.409)	(0.089)
Pval: T(Low)=C(High)	0.7265	0.3069	0.4852	0.8630	0.2915	0.1287
Pval: T(Low)=T(High)	0.9891	0.3898	0.7161	0.5449	0.5116	0.0695
Pval: C(High)=T(High)	0.4666	0.7832	0.4904	0.3473	0.3843	0.5689
Control Mean	11.504	29564.188	95816.595	64999.259	7922.917	9.741
Observations	1403	1386	1386	1408	1386	1323
Adj R-Squared	-0.001	-0.001	-0.001	0.002	0.001	0.001

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for productivity outcomes. Panel A reports ITT effects as a binary variable and spillover effects as a continuous measure set to the saturation level of each tehsil. Panel B reports saturated regression with the reference category of control units in low saturation tehsils. Panel B also reports t-tests of equality for each saturation by treatment pair.

Table 9: Take-up and Forecast Accuracy

	Weather Message Take-up
(1) Satellite Rainfall Accuracy	0.288**
	(0.119)
(2) Farmer-Reported Rainfall Accuracy	0.155*
	(0.087)
(3) Satellite Absolute Rainfall Deviation	0.025*
	(0.013)
(4) Rainfall False Alarms	-0.001
	(0.003)
(5) Rainfall RMSE	0.004
	(0.005)
(6) Max Temp RMSE	0.041**
	(0.017)
(7) Min Temp RMSE	0.013
	(0.015)
Take-up Mean	0.217
Observations	4964

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports separate regressions for the relationship between measures of accuracy (rows 1-7) and average take-up rates of weather forecast messages. The unit of analysis is tehsil by day. There are 35 unique tehsils and each received between 140-145 weather messages throughout the season. (1) refers to the share of days where rainfall forecasts aligned with realized rainfall according to satellite measures. (2) is the share of days when whether forecasts aligned with farmer-reported rainfall during the survey. (3) is the average absolute deviation of forecasts from realized weather, measured in millimeters. (4) is the number of days with false alarms in precipitation, defined as days where forecasts predicted no rain and a heavy rain fell, or when forecasts predicted heavy rain and no rain fell. (5), (6), and (7) are the root mean-squared error for rainfall, max temp, and min temp from a tehsil-level regression of the realized value on the forecast value.

Appendix

Table 10: Response Rates for Maximum and Minimum Temperature Weather Predictions

	Maximum Temperature				Minimum Temperature			
	(1) Responds to Max Temp Forecast (0/1)	(2) Responds to Max Temp Forecast (0/1)	(3) Absolute Deviation from Forecast	(4) Absolute Deviation from Forecast	(5) Responds to Min Temp Forecast (0/1)	(6) Responds to Min Temp Forecast (0/1)	(7) Absolute Deviation from Forecast	(8) Absolute Deviation from Forecast
Treat	0.046***		-1.507***		0.043***		0.019	
	(0.013)		(0.046)		(0.013)		(0.047)	
Tehsil Share Treated	-0.045		-0.212**		-0.051		-0.271***	
	(0.034)		(0.103)		(0.036)		(0.081)	
Treat in Low		0.012		-1.442***		0.005		-0.078
		(0.028)		(0.106)		(0.026)		(0.127)
Control in High		-0.014		-0.040		-0.017		-0.172**
		(0.029)		(0.107)		(0.029)		(0.083)
Treat in High		0.036		-1.570***		0.031		-0.140*
		(0.026)		(0.107)		(0.028)		(0.070)
Pval: T(Low)=C(High)		0.2632		0.0000		0.3395		0.3612
Pval: Treat(Low)=T(High)		0.2822		0.0787		0.2717		0.5412
Pval: C(High)=T(High)		0.0036		0.0000		0.0035		0.5428
Control Mean	0.438	0.432	4.056	4.227	0.438	0.432	4.915	5.446
Observations	5144	5144	5144	5144	5144	5144	5144	5144
Adj R-Squared	0.003	0.002	0.192	0.192	0.003	0.002	0.001	0.000

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT treatment effects regressions for robustness on weather predictions. Columns 1-2 and 5-6 report whether a respondent answered the survey question that asked for a prediction for maximum and minimum temperature for the following day. Columns 1 and 5 show that treated farmers are more likely to provide temperature predictions, but results are noisier and smaller in magnitude in columns 2 and 6, showing that control farmers in high saturation tehsils had similar response rates. Columns 3-4 and 7-8 report prediction errors for maximum and minimum temperature after imputing group means for non-response.

Table 11: ITT and Spillover Effects on Farming Behavior: Avoiding Forecasted Rain

Panel A: ITT and Spillover Effects as Continuous Measure

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Rain Index
Treat	-0.037	0.051**	0.025	0.035**	-0.003	-0.004	-0.004
	(0.089)	(0.020)	(0.017)	(0.014)	(0.027)	(0.016)	(0.014)
Tehsil Share Treated	0.064	0.223***	0.108**	0.150***	0.079	0.094*	0.052
	(0.196)	(0.069)	(0.046)	(0.054)	(0.066)	(0.053)	(0.034)
Control Mean	0.628	0.275	0.176	0.225	0.187	0.167	0.000
Observations	137	3213	3082	2932	968	1567	3482
Adj R-Squared	-0.013	0.017	0.005	0.009	0.000	0.002	0.000

Panel B: Spillover Effects as Saturated Regression

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Rain Index
Treat×Low Saturation	0.016	0.053**	0.027	0.008	-0.033	-0.012	-0.005
	(0.105)	(0.025)	(0.016)	(0.020)	(0.031)	(0.018)	(0.016)
Control×High Saturation	0.144	0.087**	0.046*	0.054*	-0.025	0.020	0.055
	(0.100)	(0.034)	(0.024)	(0.031)	(0.039)	(0.032)	(0.039)
Treat×High Saturation	-0.106	0.140***	0.070***	0.116***	0.021	0.033	0.051**
	(0.116)	(0.029)	(0.024)	(0.030)	(0.028)	(0.024)	(0.023)
Pval: T(Low)=C(High)	0.3191	0.3759	0.5262	0.1030	0.8399	0.3177	0.1270
Pval: T(Low)=T(High)	0.3865	0.0060	0.0916	0.0003	0.0503	0.1301	0.0252
Pval: C(High)=T(High)	0.0488	0.0676	0.4089	0.0005	0.2675	0.7172	0.9300
Control Mean	0.606	0.247	0.163	0.206	0.196	0.162	0.000
Observations	137	3213	3082	2932	968	1567	3482
Adj R-Squared	-0.010	0.012	0.004	0.011	-0.000	-0.000	0.002

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent carried out farming tasks on days without rain according to forecasts provided to farmers. Panel A reports ITT effects as a binary variable and spillover effects as a continuous measure set to the saturation level of each tehsil. Panel B reports saturated regression with the reference category of control units in low saturation tehsils. Panel B also reports t-tests of equality for each saturation by treatment pair. Sample sizes differ because responses are conditional on having reported a date for each task.

Table 12: ITT and Spillover Effects on Farming Behavior: Avoiding Forecasted Heat

Panel A: ITT and Spillover Effects as Continuous Measure

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Heat Index
Treat	-0.005	-0.005	0.012	0.018	0.041	0.059**	0.010
	(0.004)	(0.018)	(0.014)	(0.014)	(0.033)	(0.022)	(0.019)
Tehsil Share Treated	0.005	0.081	0.135**	0.004	0.244***	0.130	0.059
	(0.012)	(0.082)	(0.065)	(0.061)	(0.084)	(0.111)	(0.060)
Control Mean	0.013	0.725	0.741	0.800	0.658	0.564	0.000
Observations	2293	3303	3225	3046	1170	2081	3482
Adj R-Squared	-0.000	0.001	0.005	-0.000	0.016	0.006	0.001

Panel B: Spillover Effects as Saturated Regression

	(1) Planting	(2) Irrigate	(3) Fertilize	(4) Pesticide	(5) Fungicide	(6) Herbicide	(7) Avoid Heat Index
Treat×Low Saturation	-0.007	0.009	0.011	0.057***	0.038	0.065***	0.010
	(0.005)	(0.021)	(0.018)	(0.016)	(0.045)	(0.024)	(0.022)
Control×High Saturation	-0.006	0.062*	0.042	0.027	0.112**	0.032	0.041
	(0.006)	(0.031)	(0.029)	(0.028)	(0.049)	(0.058)	(0.043)
Treat×High Saturation	-0.000	0.034	0.062*	-0.000	0.164***	0.084	0.053
	(0.010)	(0.037)	(0.031)	(0.034)	(0.040)	(0.053)	(0.048)
Pval: T(Low)=C(High)	0.8801	0.1332	0.3403	0.1948	0.2149	0.5879	0.4946
Pval: T(Low)=T(High)	0.5241	0.5112	0.1064	0.0692	0.0108	0.7228	0.3844
Pval: Cont(High)=T(High)	0.4986	0.2515	0.3637	0.1903	0.1529	0.1747	0.6531
Control Mean	0.014	0.705	0.729	0.790	0.621	0.556	0.000
Observations	2293	3303	3225	3046	1170	2081	3482
Adj R-Squared	-0.000	0.002	0.002	0.003	0.017	0.003	0.001

Robust standard errors in parenthesis clustered at the tehsil level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table reports ITT and spillover treatment effects regressions for whether a respondent carried out farming tasks on days without extreme heat above 37C according to forecasts sent to mobile phones. Panel A reports ITT effects as a binary variable and spillover effects as a continuous measure set to the saturation level of each tehsil. Panel B reports saturated regression with the reference category of control units in low saturation tehsils. Panel B also reports t-tests of equality for each saturation by treatment pair. Sample sizes differ because responses are conditional on having reported a date for each task.