Access to Digital Extension Services and Crop Yield: Evidence from Paddy Cultivation in India*

Chanchal Pramanik$^{\psi}$

This version: November 29, 2025

Abstract

We examine the influence of information access on seasonal paddy yields in ten selected Indian States using decade-long (2009-2019) data from the Government of India's *Kisan* (farmers) call center (KCC) initiative. Based on 17 million farmers' queries at the KCC, we isolate two most frequent concerns- one related to pest pressure and the other related to weather predictions. Leveraging these farmers' queries and crop yield-related datasets at the district level, we assess the impacts of weather and paddy pest infestation information on seasonal paddy yields. Endogeneity concerns are addressed using an instrumental variable (IV) approach - previous years' paddy yields and percentage gross irrigated area as IVs for pest and weather-related queries, respectively. After confirming the validity of the IVs, the results reveal that an increase in seasonal pest attacks - proxied by a 1% increase in seasonal pest-related queries from districts - decreases paddy yields by 0.028 t ha^{-1} (i.e., a decrease in 1.1% of average paddy yields). On the contrary, a 1% increase in district-level seasonal weather queries increases paddy yields by 0.008 t ha^{-1} (i.e., an increase of 0.3% of average paddy yields). Our results point to the positive contribution, paddy yield gains, while addressing proactive types of queries (i.e., weather) as well as the limitations of digital extension services, unable to recover the potential yield while addressing reactive types of queries (i.e., paddy pests) in a developing country.

Keywords: Crop yields, Digital agriculture, Farmers' call center, Instrumental variables, Mobile phones, Pest pressure, Weather pattern

JEL Codes: Q16, R11, D83

*I am thankful to Arnab Basu, Prabhu Pingali, John Carruthers, Martin Ravallion†, Dilip Mookherjee, Natasha Jha, Phuoc Thien Anh Nguyen, Tsenguunjav Byambasuren, Patrick Hatzenbuehler, and seminar participants at Cornell University, the Ninth International Conference on Agricultural Statistics (ICAS IX), Washington D.C. (2023) and International Association for Applied Econometrics (IAAE), Turin, Italy (2025) for their valuable comments and feedback.

$^{\psi}$Post-doctoral Fellow in Agricultural Economics, University of Idaho and Research Associate, Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, New York, USA, Email: cp629@cornell.edu

1 Introduction

The uptake of digital technologies, especially mobile phones, by the rural population in developing countries has allowed for a rapid expansion of digital extension services in agriculture.[1] Various studies have shown that digital extension services in the form of advanced information on weather, input utilization, pest management strategies, and market price fluctuations can significantly improve agricultural planning and operational efficiency for small and marginal farmers (Aker, 2011; Davis and Sulaiman V, 2014; Haile et al., 2016; Letta et al., 2022; Gupta et al., 2024). It is further argued that direct access to information from experts via mobile phones can be more efficient in enhancing productivity than learning from peers (Casaburi et al., 2014; Kondylis et al., 2017; Cole and Fernando, 2021).[2] For instance, the Government of India launched the *Kisan* (farmers) call center (KCC) initiative in 2004 to provide advisory support to Indian farmers via mobile phones in their respective regional languages. At these KCC centers, agents with formal training in agricultural sciences respond to context-specific farming queries, offer tailored guidance, and systematically document each interaction by transcribing the conversations into English-language text records (Dominic et al., 2022; Gupta et al., 2024). The KCC services and other mobile-based services in agriculture have improved crop productivity in developing countries.

Nevertheless, there is limited empirical evidence on the impact of *particular types of information* accessed primarily via mobile phones on crop yields in developing countries. In this paper, we focus on two particular types of information (weather and pest-related) via mobile phones on the yield of an important staple crop, i.e., paddy[3], in India. We address the above research question by analyzing more than 17 million small and marginal farmers'[4] queries recorded at the KCC in ten

[1] Private and public sector investments in the expansion of electricity grids and mobile phone towers have sped up digital technology adoption in developing countries (Chaurey et al., 2004; Agrawal et al., 2020; Gupta et al., 2024).

[2] Additional studies have also shown that digital technologies have a significant impact on agricultural productivity in developing countries (Deichmann et al., 2016) and the use of mobile phones, in particular, significantly reduces price dispersion in agricultural markets in India and Niger (Jensen, 2007; Aker, 2010; Parker et al., 2016).

[3] Indian agriculture depends on the yield of paddy, one of the major staples cultivated in all cropping seasons across regions. Farmers in the country have achieved 40-50% of the yield potentials (i.e., maximum possible output from its given sets of inputs, including genetic potential of the plants) (Lybbert et al., 2023; Das, 2012). Moreover, India contributes the highest share (40% in 2022) of rice exports in the world (Nayak et al., 2024). Judicial use of inputs has the potential to narrow the yeld gaps and enhance food security for India and the world (Bandumula, 2018; Nayak et al., 2024).

[4] Small and marginal farmers (i.e., agricultural land holdings less than 2 hectares) own 82% agricultural land in India, Our World in Data.

selected States[5], covering major agroclimatic zones in India[6] in all three cropping seasons[7] from 2009 to 2019. At the KCC, Indian farmers are able to access farm-specific and real-time scientific agricultural information throughout the crop cycle, enabling them to make informed decisions and optimize their agricultural practices (Gupta et al., 2024). While farmers access a variety of information from the KCC, paddy crop-related and pest and weather issues-related queries were the most frequent[8]. A case study conducted by the Indian Institute of Management, Ahmedabad (IIM-A) on Gujarat farmers suggests that the farmers reported improved awareness of pest and weather-related crop management (Gandhi and Johnson, 2018). While researchers from both IIM-A and the Indian Council of Agricultural Research's National Academy of Agricultural Research Management (ICAR-NAARM) identified opportunities to improve the service of KCC on information content and integration of regional agricultural knowledge systems (Gandhi and Johnson, 2018; Yashavanth et al., 2022). We hypothesize that the incidences of these particular query types at KCC have a significant positive impact on seasonal paddy yields.[9] Specifically, farmer queries through mobile phones related to paddy pests are anticipated to enhance crop protection strategies and improve yield, while those concerning weather conditions are expected to address issues related to water stress in paddy cultivation and improve yield. Comprehending these causal relationships is crucial for enhancing KCC services (i.e., Indian agriculture's digital extension services[10]).

However, establishing causal relationships between incidences of pest and weather-related queries and seasonal paddy yields is challenging. Farmers' queries on paddy-related issues may be endogenous to the paddy yields at district levels due to several factors, such as simultaneity

[5]Regions, State names and percentages of KCC queries: North (Punjab: 11.04%, Haryana: 11.56%), South (Andhra Pradesh: 4.44%, Telangana: 4.40%, Tamil Nadu: 8.16%), East (Odisha: 6.39%, Bihar: 10.36%), West (Maharashtra: 20.63%, Gujarat: 7.45%), Central (Madhya Pradesh: 15.57%), Table 2.

[6]Agro-climatic zones and States: Trans-Ganga plains region (Punjab and Haryana), Middle Gangetic plains region (Bihar), Eastern plateau and hills and Coastal regions (Odisha), Central plateau and hills (Madhya Pradesh), Gujarat plains and hills (Gujarat), Western plateau and hills and coastal plains and ghats (Maharashtra), Southern plateau and hills and coastal plains (Telangana, Andhra Pradesh and Tamil Nadu).

[7]Seasons and percentages of KCC queries: Monsoon (*Kharif*): 54%, Winter (*Rabi*): 29% and Summer: 17%.

[8]In our study, paddy-related queries constitute 8% of total queries, highest among food grain and individual crop types. While pests and weather-related queries constitute 33% and 35% of total queries, Table 1. Crop pests are pathogens and insects that cause economic damage to crops, following the USDA definition of Integrated Pest Management (IPM).

[9]The hypothesis aligns with the results obtained from the OLS estimates as shown in Table 6, particularly in columns 1 to 3.

[10]Creation of agricultural information involves fixed costs, but once created, it can be used by more people at a minimal marginal distribution cost, with no cost to others. Thus, research initiatives aimed at improving digital agricultural extension services (such as, KCC in India) in developing nations should focus on strengthening their respective public infrastructures (Fabregas et al., 2022).

and omitted variable biases. In terms of simultaneity bias in our analysis, farmers can be caught in a cycle where resource-rich farmers in historically productive areas have the resources (e.g., higher quality smartphones) to seek advanced information from KCC compared to farmers in less productive areas who are resource-poor. As a result, one might observe a higher volume of queries from resource-rich productive areas. In terms of omitted variable bias, additional unobserved characteristics that lead to the incidences of queries related to pests and weather, although captured in the error term, nevertheless confound with the incidences of particular queries for paddy. To address the issue of endogeneity between seasonal paddy yields and the incidence of query types in the KCC, we employ an instrumental variable (IV) approach along with fixed effects at district and season levels. We consider seasonal paddy yields from previous years and the percentage of gross irrigated area in the districts as instrumental variables for the percentage of seasonal queries related to pests and weather in each district, respectively[11]. To formally account for the district-level heterogeneity, as part of robustness checks, we introduce interaction terms between query incidence and district-level resource indicators (e.g., paddy price shocks, night light luminosity, agro-climatic zones, average landholding size, and gender wise agricultural landholding, mobile ownership, and literacy rates) in our regression framework. This allows us to test whether the effect of queries on yield varies systematically across districts with differing resource endowments.

Our causal identification depends on the selection and validity of the instrumental variables. Paddy pests result in paddy yield losses in the range of 37.4% to 51.4% throughout the world and in the range of 21% to 51% in India (i.e., yield losses of paddy are assessed by comparing them to the optimal yield achieved in agronomic trials, where crops are grown free from pest infestations). (Oerke, 2006; Sharma et al., 2017). It is reasonable to assume that the losses related to pests in the previous year should be highly correlated with the pest-related queries in the observed year. On the other hand, paddy being a water-sensitive crop, weather-related queries (concerned primarily with rainfall forecasts) allow farmers to schedule irrigation in their fields. These observations guide our choice of the two instruments for the endogenous regressors. Rigorous tests

[11] As our study utilizes panel agricultural data, we examined the presence of serial correlation within the outcome and treatment variables (Tsay, 2005). The autocorrelation function plots at the state level indicate there is no significant serial correlation (Figure 5). Further, we conducted formal LM, Wooldridge, and Durbin-Watson tests for panel AR(1) (Table B.2) followed by Driscoll-Kraay or PCSE robustness (Table 7). Moreover, we perform a regress of the instruments against all the control variables. Should the instruments display a strong correlation with the observables, it is probable they are also correlated with the unobservables (Altonji et al., 2011). Most of the control variables are not significantly correlated, and a few control variables (i.e., average landholding size, fertilizer consumption, and price shocks) are weakly correlated with the instruments, which validates the appropriateness of the selected instruments (Table B.9).

to validate these instruments, including assessments of endogeneity (i.e., OLS inconsistent if regressors are endogenous, but 2SLS is consistent) and relevance (i.e., instruments correlated with regressors), reconfirm our choice of instruments. Additionally, we test the exclusion restriction assumption (i.e., instruments affect outcome only through the regressors) using an informative zero-first-stage test (Altonji et al., 2005; Angrist et al., 2010; Van Kippersluis and Rietveld, 2018). This test verifies that the instrument has no direct effect on the outcome in a sub-sample[12] where it has no first-stage effect on the treatment variable, as detailed in later sections (section 4 and subsection 5.2). Also, we test the direction of the first-stage regression estimates for the IVs for different sample groups, such as regions and States, to validate the exclusion restriction and monotonicity assumption (Imbens and Angrist, 1994; Angrist and Pischke, 2009). We further analyze whether the instruments are associated with other yield factors, such as total fertilizer consumption per gross cropped area, rainfall, and price shocks. The instruments are irrelevant to these outcomes, providing further evidence supporting the exclusion restriction (more details in Appendix subsection B). Our IV estimates of the two most frequent queries show significant opposing effects on seasonal paddy yields. For instance, a higher frequency of pest-related queries triggered due to higher incidences of paddy pests (such as a reactive approach for mitigation and damage control) in fields negatively affects paddy yields, whereas a higher frequency of weather-related queries (such as a proactive approach for prevention and readiness) has a positive impact on the same [13]. This inference falsified our hypothesis related to pest queries.

An intuitive explanation of our findings is that paddy requires a wet environment, which also

[12]Selection criteria of sub-samples: Farmers' queries from the eastern (Bihar and Odisha) and western (Gujarat and Maharashtra) regions are predominantly focused on schemes, financial support, and contact information to access the government benefits. Together, these query types account for approximately 28% of the total queries, ranking second only to weather-related queries in the respective regions (Figure 1). Besides the nature of queries, in the western region, such as Gujarat, farmers focus on wheat cultivation (Gulati et al., 2021), and Maharashtra lacks irrigation support (Appendix subsection C). The incidence of these particular queries does not significantly affect paddy yields in the districts of these regions (i.e., we have found an insignificant correlation between instrument variables, treatment variables, and the outcome variable). The zero first-stage samples are selected on the basis of the individual instrument variables. The districts of Bihar, Odisha, Gujarat, Haryana, and Punjab are selected for the previous years' paddy yields, which is the IV for pest-related queries. For the percentage of gross irrigated area (i.e., IV for weather-related queries), districts of Bihar, Maharashtra, and Punjab are selected. The districts of Punjab and Haryana have reached higher yield potentials (i.e., maximum possible output from their given sets of inputs, including the genetic potential of the plants, following the definition of technical efficiency) for paddy, which could mean that KCC queries may not significantly enhance further yield. However, access to KCC information is crucial for sustaining these higher yields in the northern Indian States.

[13]Given that the farmers' distress calls to KCC stemmed from pest infestations evidenced in paddy fields, recuperating potential yields proved challenging. Establishing preventive crop pest control infrastructures in the regions, informed by historical crop pest query trends and yield reduction data, would benefit the farmers. Conversely, weather-related queries about rainfall and forecasts serve a preventive role, aiding in paddy crop management and planning, thereby enhancing yield outcomes.

favors the growth of insects and pathogens. Given that farmer calls usually happen after visual evidence of diseases and insect infestations, it is perhaps not surprising that the direct negative effect of the pest attack is only partly ameliorated by the positive effect of query conditional on the attack. On the other hand, farmers who rely on weather-related data can effectively plan their agricultural tasks, such as transplanting, irrigation timing, and harvesting. These practices support plant growth and results in higher seasonal paddy yields when compared to those farmers lacking relevant weather insights (Fabregas et al., 2019; Gupta et al., 2024). Therefore, it is safe to conclude that the positive impact of digital technologies on crop yield is via queries that allow farmers to plan in advance.

The remainder of the paper is organized as follows, section 2 provides an overview of the data sets and variables involved in the analysis; section 3 explores the empirical strategy; section 4 discusses the identification strategy; section 5 summarizes the empirical results followed by robustness checks in section 6. Finally, section 7 concludes the paper.

2 Data and Measurement

Our empirical analysis is based on multiple data sources. The first is a dataset that contains more than 17 million farmers' queries recorded at the KCC from across India. We focus on the queries from ten States[14] representing the main geographical and agroclimatic regions over the 2009-2019 (pre-COVID-19 lockdowns in early 2020). The second dataset contains weather variables, including maximum and minimum temperature, and rainfall, which we access at the month-by-district levels from 1999 to 2019. The additional ten years (1999-2008) of the weather variables are used to construct historical averages, which are used to construct deviations over the period of our study, 2009-2019. The third dataset, sourced from multiple sources related to seasonal paddy yields[15] at the district level, such as percentage of rural population, total agricultural land, input usage (gross irrigated area and total fertilizer consumption), and infrastructure (urban local bodies, markets, road length, and railways), are used for our analysis.[16] Lastly, the fourth dataset is composed of a

[14]Regions and State names: North (Punjab, Haryana), South (Andhra Pradesh, Telangana, Tamil Nadu), East (Odisha, Bihar), West (Maharashtra, Gujarat), Central (Madhya Pradesh).

[15]Besides paddy yields, we access seasonal wheat and millets yields to present placebo checks of our analysis.

[16]The weather variables, amount of agricultural land, fertilizer consumption, and percentage of rural population are used as baseline control (i.e., biophysical environment) while variables related to the infrastructure supporting agriculture in the districts are used as additional controls. A detailed list of variables is described in Table 3.

set of indicators - specifically, agricultural land ownership, literacy rates and mobile phone ownership disaggregated by gender, night-time light intensity, and shocks to paddy prices - compiled from several independent data sources. These estimated variables are incorporated into the study to conduct robustness checks and thereby assess the stability and reliability of the main empirical findings. In the subsequent sections, we present the data description and construction of the variables for analysis.

2.1 Incidence of Queries at *Kisan* (Farmers) Call Center

The Ministry of Agriculture and Farmers Welfare, Government of India, introduced the KCC initiative in 2004, recognizing gaps in agricultural extension services. KCC facilitates telephonic assistance to farmers in 22 vernacular languages with the support of trained teleadvisors[17] subsection A. covering various aspects of agriculture. The rapid and extensive penetration of basic mobile phones across developing regions creates significant potential for initiatives such as KCC to complement and enhance traditional, facetoface agricultural extension systems. These conventional modalities are often characterized by high operational costs, logistical inefficiencies, and persistent challenges in monitoring performance and ensuring accountability, all of which mobile phone-based services are well-positioned to mitigate and partially reform (Fabregas et al., 2019). The KCC dataset offers textual data on the conversations between farmers and call center agents based on farmers' locations and specific times of calls. These textual conversations are processed through text mining and classified under several English, non-English (i.e., local agronomic terms), and technical words. These word classifications form the basis of developing an agronomic data dictionary based on crops and issues, and categorizing the conversations. For our analysis, we have categorized farmers' queries into different types based on dummy variables that represent various issues corresponding to the main crops (i.e., paddy, wheat, pulses, millets, vegetables, etc.) and issues. Key agricultural issues[18] include weather, pests, finance, prices, and inputs. Among these different types of queries, weather and pest-related ones have the highest average frequencies. Similarly, queries related to paddy and wheat show the highest averages

[17]The farmers' queries and answers from KCC are recorded as texts in English in the KCC dataset. More details of the KCC initiative are in Appendix

[18]Finance: Queries are related to direct benefit transfer, *Kisan* (farmer) credit card, loan, subsidies, market and schemes; Price: Queries related to crop prices, minimum support price; Inputs: Queries related to seed varieties, cropping practices, and irrigation; Contact: Queries related to contact information to Krishi Vigyan Kendra (agricultural extension center), state government

among individual food crops, while queries related to vegetables[19], pulses, millets, maize, sugarcane, oilseed, fodder, cotton, and livestock are also observed, albeit at lower frequencies. Detailed summary statistics of the query types are in Table 1.

The distributions of query types at the regional level are plotted for paddy in Figure 1. Almost half of the queries are related to the weather, followed by paddy pests. Regarding paddy pests, the exception is the western region (which includes the States of Gujarat and Maharashtra), which experiences most of the weather deviations (Figure 2) and is not known for significant paddy cultivation. Table 2 provides information on the State-wise distribution of queries, including those related to pest and weather. Tamil Nadu, Gujarat, and Odisha record fewer weather-related queries compared to other States. Districts of Tamil Nadu and Odisha cultivate paddy across all the seasons (more detailed analysis on paddy yield distribution is in Appendix subsection C) but had recorded fewer rainfall deviations (Figure 2). On the other hand, Gujarat farmers are not focused on paddy cultivation. Hence, there were differences in the distribution of weather-related queries at KCC. Overall, majority of the queries are related to weather ($\sim$ 35%) and are mostly recorded from the northern States of Punjab and Haryana. While comparing pest-related queries, Bihar recorded fewer queries than other States, though Bihar cultivates paddy across all paddy-growing seasons (Appendix subsection C). Further details on the KCC dataset can be found in the Appendix subsection A.

2.2 Weather Variables Influencing Crop Yields

We obtain weather data from the TerraClimate dataset[20], which provides monthly maximum and minimum temperatures and rainfall data. The weather variables reported correspond to the average of the daily maximum and minimum temperatures within a month at the district level, while rainfall represents the total district-level rainfall recorded over the month. Weather variables are aggregated at the seasonal level for analytical purposes[21]. The dataset covers each district within

[19]The vegetable group contains multiple crops; we do not have yield data for every type of vegetable discussed at KCC for each district. The mean queries of the vegetable group is 13%.

[20]The TerraClimate data is processed by analyzing spatial data in NetCDF obtained from Climatology Lab. ICRISAT's district-level weather data are also processed from this data repository.

[21]The weather data for each district on a monthly basis are associated with their corresponding seasons. The maximum temperature of a season in a district is determined by the maximum of the monthly maximum temperatures within that season. Similarly, the minimum temperature of a season in a district is ascertained from the minimum of the monthly minimum temperatures in that season. The seasonal rainfall in a district is calculated by summing up the monthly rainfall totals within the season.

the ten selected States and the period of 21 years (1999-2019). An additional 10 years of weather data (1999-2008) are included to design historical weather deviation variables that include the maximum temperature deviation ($maxTD_{(ds)ry}$), minimum temperature deviation ($minTD_{(ds)ry}$), and rainfall deviation ($RD_{(ds)ry}$)[22]. In the analysis, the indicator variables capturing weather deviationsconstructed from historical weather recordsare defined as relative deviations from typical conditions, thereby representing variations in weather patterns that could plausibly have influenced crop yields. The equations for the weather deviation variables are the following.

$$maxTD_{(ds)ry} = \begin{cases} 0 & \text{if } \frac{1}{21}\sum_{y=1999}^{2019} maxT_{(ds)ry} - 2\sigma(maxT_{(ds)r}) \leq maxT_{(ds)ry} \\ & \leq \frac{1}{21}\sum_{y=1999}^{2019} maxT_{(ds)ry} + 2\sigma(maxT_{(ds)r}), \\ 1 & \text{otherwise.} \end{cases}$$

$$minTD_{(ds)ry} = \begin{cases} 0 & \text{if } \frac{1}{21}\sum_{y=1999}^{2019} minT_{(ds)ry} - 2\sigma(minT_{(ds)r}) \leq minT_{(ds)ry} \\ & \leq \frac{1}{21}\sum_{y=1999}^{2019} minT_{(ds)ry} + 2\sigma(minT_{(ds)r}), \\ 1 & \text{otherwise.} \end{cases}$$

$$RD_{(ds)ry} = \begin{cases} 0 & \text{if } \frac{1}{21}\sum_{y=1999}^{2019} R_{(ds)ry} - 2\sigma(R_{(ds)r}) \leq R_{(ds)ry} \\ & \leq \frac{1}{21}\sum_{y=1999}^{2019} R_{(ds)ry} + 2\sigma(R_{(ds)r}), \\ 1 & \text{otherwise.} \end{cases}$$

where, $\sigma(maxT_{(ds)r})$, $\sigma(minT_{(ds)r})$ and $\sigma(R_{(ds)r})$ denote the standard deviations of maximum temperature, minimum temperature, and rainfall, respectively.

The heat maps illustrated in Figure 2 (I, II and III) show the variations in maximum and minimum temperatures, along with rainfall across districts in the chosen States of India. The districts of Maharashtra, Madhya Pradesh, Odisha, and Tamil Nadu present higher deviations in maximum temperatures compared to other States. The districts of Bihar and Tamil Nadu display significant variations in minimum temperatures. Regarding rainfall deviations, Madhya Pradesh, Gujarat, and Odisha are noteworthy. The average seasonal deviations in maximum and minimum temperatures and rainfall among the 314 districts in the selected States during the period from 2009 to 2019 are 2.4, 2.1 and 3.0, Table 3.

[22]The subscripts denote the following: d stands for District, s stands for State, r stands for Season, and y stands for Year.

2.3 Rural population, agricultural land, irrigation coverage, fertilizer consumption, infrastructure, and crop yields

Rural population, agricultural land, fertilizer consumption, infrastructure, and paddy prices influence paddy yields at the district level, along with KCC queries and weather factors. The discussion on these variables are as follows:

2.3.1 Rural Population

Rural population not only provides agricultural laborers but also influences demand for food grains (Emerick, 2018; Komarek and Msangi, 2019; Krishna and Meenakshi, 2022). We include the district-wise percentage of the rural population in our analysis. The population census database and estimates from Ministry of Statistics and Programme Implementation's (MOSPI) statistical year books contribute to the district-level rural and State-level projected population data. The cohort component method is the primary and most detailed method, factoring in fertility, mortality, and migration to project population by age and sex cohorts. MOSPI used this technique for national and large state-level projections, supported by data from the population census, the Sample Registration System, and health surveys (Office of the Registrar General & Census Commissioner, 2020). The State-level population projections from MOSPI are distributed to the districts based on estimated district-level population shares derived from census data (United Nations Population Fund, India, 2009). We use these estimates to evaluate the percentage of the rural population at the district level using the following equation. Rural population is denser in the districts of Bihar, Odisha, Madhya Pradesh, and Telangana States Figure 2 (IV). The average district level rural population is 78.21%, Table 3.

$$\%\ \text{Rural Population}_{(ds)y} = \frac{\text{Rural Population}_{(ds)y}}{\text{Total Population}_{(ds)y}}$$

2.3.2 Agricultural Land

District-level gross cropped area for paddy, and total district area are accessed through Department of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Government of India. India's agricultural census database is referred to for accessing data on the distribution of

agricultural land holding sizes and the average agricultural land holding size in the districts. The following agricultural land-related variables are developed for our analysis. The percentage of gross cropped area captures the availability of the agricultural area, and the Gini coefficient captures the inequality of agricultural land in a district. These agricultural land variables reflect yield variability across the districts.

$$\%\ \text{Gross Cropped Area (GCA)}_{(ds)y} = \frac{\text{Gross Cropped Area}_{(ds)y}}{\text{Total district area}_{(ds)y}}$$

$$\text{Gini coefficient (land holding size)}_{(ds)y} = \frac{\sum_{i=1}^{n} n_i \sum_{j=1}^{n} n_j |L_i - L_j|}{2(\sum_{i=1}^{n} n_i)^2 \bar{L}}$$

where, n_i and n_j are the number of land holdings of sizes L_i and L_j, and $\bar{L}$ is the average land holding size.

The summary statistics of these agricultural land variables are presented in Table 3. The average gross cropped area, agricultural land holdings, and Gini coefficient (land holding size) are 93%, 1.47 hectares, and 0.47, respectively. The heat map shown in Figure 3 (I and II) provides an overview of agricultural land availability and its inequality across the selected States. It reveals that above-average gross cropped areas for paddy are in the districts of Punjab and Haryana, and part of Madhya Pradesh, Maharashtra, and Bihar. While based on the land inequality measure, land sizes are more equal in the districts of Punjab, Odisha, Gujarat, and Maharashtra, whereas most land size inequality is in the districts of Haryana.

2.3.3 Input Variables

The input variables (i.e., gross irrigated area and total fertilizer consumption) are accessed through District-Level Database (DLD) of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). The following input variables are developed for our analysis.

$$\%\ \text{Gross Irrigated Area}_{(ds)y} = \frac{\text{Gross Irrigated Area}_{(ds)y}}{\text{Gross Cropped Area}_{(ds)y}}$$

$$\text{NPK per GCA}_{(ds)y} = \frac{\text{Total fertilizer consumption (NPK)}_{(ds)y}}{\text{Gross Cropped Area}_{(ds)y}}$$

The Figure 3 (III and IV) depicts the state of agricultural input consumption, specifically fertilizer and irrigation. The heat map highlights that Odisha, Madhya Pradesh, and Maharashtra have recorded below-average irrigation water consumption and total fertilizer use. The above-average consumption of these inputs is in the districts of Punjab, Haryana, Bihar, Telangana, Andhra Pradesh, and Tamil Nadu. The average district-level gross irrigated area and fertilizer consumption per gross cropped area are 0.51% and 156 kg ha^{-1}, Table 3.

2.3.4 Infrastructure Variables

Local infrastructures influence crop yield growth by facilitating market access in developing countries (Zhang and Fan, 2004; Shively and Thapa, 2017). The presence of adequate infrastructure facilities in districts has a significant impact on crop yield. The transportation system, which comprises roads and railways, plays a crucial role in providing access to markets and catering to the urban food demand. Districts with well-developed transportation networks, enabling easy connectivity to urban areas, tend to experience higher agricultural productivity (Binswanger et al., 1993; Fan et al., 2000). The empirical analysis in this study incorporates several infrastructure variables, including the length of the roads (i.e., State and national highways and district/village roads maintained by local bodies) and the number of railway stations per district area, the number of markets (i.e., regulated and submarkets), and the presence of urban local bodies such as municipal corporations, municipalities, and notified area council within the districts. These infrastructure data on roads and markets are accessed through ICRISAT's District-level Data for India, while urban local bodies data are accessed through individual district websites following Integrated Online Government Directory, and district-wise railway stations list are obtained through Ministry of Railways.

The equations of the infrastructure variables are as follows:

$$\text{Road length per total dist. area}_{(ds)y} = \frac{\text{Road Length}_{(ds)y}}{\text{Total district area}_{(ds)y}}$$

$$\text{Train stations per total dist. area}_{(ds)y} = \frac{\text{Number of train stations}_{(ds)y}}{\text{Total district area}_{(ds)y}}$$

The heatmaps in Figure 4 (I to IV) show the distribution of roads, railways, markets, and urban local bodies across the districts of the selected States in India. Odisha and Tamil Nadu have more road networks per unit of district area, while more railway stations are in parts of Bihar, Punjab, Gujarat, and Tamil Nadu. In comparison to the number of markets, Maharashtra, Punjab, Odisha, and Bihar are the leading States. On the other hand, Tamil Nadu and the western part of Maharashtra have more urban local bodies. The district-level average of road length, number of railway stations, markets, and urban local bodies are 10.42 km ha^{-1}, 0.05 ha^{-1}, 15, and 7, respectively, Table 3.

2.3.5 Crop Yields

The agricultural outcome variable focuses on seasonal paddy yields, recorded at the district level during the monsoon (i.e., *Kharif*), winter (i.e., *Rabi*), and summer seasons. The seasonal paddy yields at the district-level are accessed through Department of Economics and Statistics (DES), Ministry of Agriculture and Farmers Welfare, Government of India. The average district-level seasonal paddy yields is 2.60 t ha^{-1}, Table 3. It is important to note that the months of the agricultural seasons vary across different States and regions in India (Table 4). The seasonal yield data for the crops are appropriately matched with the corresponding months for each State in the database. Similarly, we integrate the wheat and pulses[23] seasonal yields data with the analysis dataset for placebo checks in our analysis. The average seasonal yields for wheat and pulses are 2.96 and 0.70 t ha^{-1}, respectively (Table 5).

District-level yield estimates from the DES are the most comprehensive official source available but may contain non-classical measurement errors due to underreporting or smoothing of yields, particularly during years of poor harvests. Such errors would generally attenuate the estimated effects of information access through KCC toward zero, implying that our results likely represent lower-bound estimates of the true effects. While we cannot directly validate yield accuracy at the district level, but the paddy price shocks are infrequent and vary modestly across years Fig-

[23]Pulses group includes the following crops: Pigeon Pea, Cowpea, Green Gram, Black Gram, Lentil, Grass Pea, Moth Bean, Peas and Beans, Other monsoon pulses, Other winter pulses, Other summer pulses.

ure 7 (IV), and the high correlation with independent sources such as national production trends provides reassurance that our estimates capture genuine variation in agricultural performance.

In the Appendix subsection C section, there is a detailed discussion on the paddy yield variation based on seasons Figure C.1 to Figure C.3. Monsoon is the main paddy growing season, followed by winter and summer. Northern and southern Indian States are the centers for paddy-growing hotspots or high-yielding district clusters. The eastern, western, and central Indian districts form the low-yielding district clusters.

2.4 Other Variables for Robustness Checks

To perform robustness analysis, we include additional variables from multiple sources. These variables include calculating weather z-scores, estimating agricultural land and mobile phone ownership and literacy rate based on gender, night light luminosity reflecting economic progress, and paddy price shocks. The descriptions of these variables are presented in the following subsections.

2.4.1 Weather variables z-scores

As an alternate measure of the weather deviations, we include the z-score of the weather variables in the analysis. The z-scores of the weather variables are calculated and integrated with the analysis databases based on the following formula. The spatial distributions of the z-scores are similar to those of the weather deviation variables Figure 2 (I to III) and the summary statistics are presented under Table 5.

$$z - ScoreW_{(ds)ry} = \frac{W_{(ds)ry} - \mu W_{(ds)r}}{\sigma W_{(ds)r}}$$

where, (ds) is district-State observation, r denotes season, y denotes year and W represents observed seasonal weather variables (maximum temperature, minimum temperature, and rainfall), μW is the mean of weather variables based on seasons and districts, and σW represents the standard deviation of weather variables based on seasons and districts.

2.4.2 Demographic and Health Survey Variables: Agricultural Landholding and Literacy Rate, Mobile Phone Ownership

Further, we utilize extensive data from the Demographic and Health Survey (DHS) in India. This comprehensive survey provides detailed information at both the individual and household levels, encompassing various facets of health and demographic characteristics. Additionally, it offers insights into agricultural landholdings, literacy rates, and mobile phone ownership, with all data disaggregated by gender. The DHS-4 (2015-16) and DHS-5 (2019-21) datasets contain mobile ownership information. We estimated the percentages of agricultural landholding, literacy rate, and mobile ownership by male and female adult populations in rural areas at the district level. These estimated variables for DHS-4 are distributed to the years 2009 to 2015, and DHS-5 to the years 2016 to 2019, for each district in the analysis dataset.

The distributions and summary statistics of the agricultural landholding by gender are presented in Figure 6 (I and II) and Table 5. Male agricultural land ownership across all districts averages 91%, whereas female agricultural land ownership averages around 11%. The above-average male agricultural land ownership districts are concentrated in all the selected States except Telangana and Andhra Pradesh. These two states rank at the top regarding female agricultural land ownership. Above-average literacy rates are in Maharashtra, Tamil Nadu, Punjab, Haryana, and parts of Odisha across the genders. The average male and female literacy rates are 72% and 57%, respectively. While comparing the mobile phone ownership based on males and females (Figure 6 (I and II) and Table 5), the averages are 92% and 91%, respectively. Except for Odisha and eastern Madhya Pradesh, all districts have above-average mobile phone ownership across genders.

2.4.3 Night Light Luminosity

Besides mobile phone accessibility, we access night light luminosity data through Development Data Lab, sourced from NASA/NOAA[24] Visible Infrared Imaging Radiometer Suite (VIIRS) and following the techniques demonstrated by Elvidge et al. (2021) from 2012-2019. We utilize Defense Meteorological Satellite Program - Operational Linescan System night light data for the remaining previous years (2009 to 2011), following the same source. Night light luminosity is frequently

[24]NASA: National Aeronautics and Space Administration, NOAA: National Oceanic and Atmospheric Administration, United States Government.

employed as an indirect measure for assessing the presence of electrification or the level of economic activity, particularly in scenarios where longitudinal data on economic performance is not readily accessible. In our approach, we utilize the metric of average night light, which is derived by aggregating the radiance values from all pertinent pixels and subsequently dividing this total by the count of pixels located within the specific district-level geographical boundary. The above-average night light luminosity is recorded in the districts of Punjab, Haryana, and parts of Tamil Nadu, Telangana, Gujarat, and Maharashtra. The low-performing districts based on night light are in Madhya Pradesh, Odisha, and Bihar (Figure 6 (III)). The average night luminosity is 2.14 nW/cm^2/sr (Table 5).

2.4.4 Paddy Price Shocks

Next, we measure the paddy price shocks variable based on the following formula to capture any shocks related to pest attack or other agronomic or economic shocks impacting paddy yield. The technique is similar to the weather deviation calculations. The yearly paddy price data at the district-level area are accessed through ICRISAT's District-level Data for India. We don't observe significant paddy price shocks in the selected districts in India (Figure 6 (IV)), with an average of 0.03 (Table 5).

$$Price\ Shock_{(ds)y} = \begin{cases} 0 & \text{if } \frac{1}{11}\sum_{y=2009}^{2019} Price_{(ds)y} - 2\sigma(Price_{(ds)}) \leq Price_{(ds)y} \\ & \leq \frac{1}{11}\sum_{y=2009}^{2019} Price_{(ds)y} + 2\sigma(Price_{(ds)}), \\ 1 & \text{otherwise.} \end{cases}$$

where, $\sigma(Price_{(ds)})$ denotes the standard deviations of paddy prices at the district-level.

3 Empirical Strategy

This section outlines the empirical strategy used to investigate the causal relationships between the incidence of farmers' queries and seasonal paddy yields in India. Categorizing the conversations between farmers and KCC agents reveals the information needs for paddy cultivation by region in India (Table 1 and Figure 1). The growth of farmers' queries over the years is signifi-

cant[25] across ten selected States (Compound Annual Growth Rate (CAGR): 31%, 2009-2019), Table A.2. The seasonal variability of the incidence of queries reflects regional priorities in cropping patterns (Table A.3). Again, queries related to pests and weather constitute 33% and 35% of the total queries, respectively (Table 2), emphasizing the increasing need for information support in these two components for paddy cultivation. But only 1.17% of the queries comprise both pests and weather-related. We aggregated these query types at the district and seasonal levels since we have seasonal paddy yields and associated datasets accessible at the district level from 2009 to 2019. The district level is the most granular level on which we can base our empirical analysis. We believe that district and seasonal data levels serve as a useful proxy for farm-level cropping decisions for three main reasons: (a) a district remains a sufficiently small unit for crop production within the grain market; (b) there is considerable variability in the incidences of KCC queries, input use, paddy yields, soil quality, and weather conditions across districts[26]; and (c) seasonal attributes related to weather conditions, paddy yields, and KCC queries also add considerable variability. The following subsections discuss the main empirical specification and the issue of endogeneity.

3.1 Empirical specifications

Between 2009 and 2019, the number of farmer queries received through the KCC platform grew considerably across the ten States, reflecting the rapid diffusion of digital information channels among farming communities. This expansion motivates our regression analysis: if increasing access to timely and relevant knowledge improves farmers responsiveness to pest outbreaks and weather risks, districts with greater intensity of information demand should experience faster yield convergence over time. Hence, the growth in query volume provides the empirical basis for testing whether information diffusion translates into measurable yield gains in the econometric specifications that follow. Further, several papers propose conceptual frameworks for how

[25] A national survey carried out in India between 2017 and 2018 discovered that 56% of farming households obtained technical guidance on paddy cultivation from extension services in the country (Kumar et al., 2021). Fellow farmers serve as reliable information disseminators in India, fostering the adoption of agricultural technologies (Foster and Rosenzweig, 1995; Emerick and Dar, 2021). Consequently, the agricultural knowledge provided by extension services is transferred to larger farming communities.

[26] Several researchers have utilized county-level data in the United States to investigate how crop yields respond to diverse factors through panel data instrumental variable approach (Miao et al., 2016; Li et al., 2019). We contend that a similar approach is applicable at the district level in India. A similar methodological study on panel data instrumental variable estimator at the district level is reported in the Indian context related to gender and farm mechanization (Afridi et al., 2023).

digital technologies (mobile phones, internet, ICT) may reduce information and transaction-cost barriers in agriculture - potentially increasing inclusion, productivity, and innovation (Table B.1). ICT impact may suffer from endogeneity (e.g., farmers with better access to resources are more likely to adopt digital tools) and omitted variable bias (other market or infrastructure constraints), which justifies using causal-identification methods such as instrumental variables or randomized designs. The main regression model used in this study is defined as follows:

$$\begin{aligned} Yield_{(ds)ry} =& \alpha + \beta_1 Queries_pest_{(ds)ry} + \beta_2 Queries_weather_{(ds)ry} \\ &+ \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} + \eta_{(ds)} + \psi_r + \eta_{(ds)} \times \psi_r + \epsilon_{(ds)ry} \end{aligned} \tag{1}$$

In equation 1, $Yield_{(ds)ry}$ represents the seasonal yields of paddy in district d, State s, season r and year y. The coefficients β_1 and β_2 capture the effect of changes in the percentage of seasonal queries related to pests and weather for the respective crops at the district level on the yields of the current year. Instrumental variables are introduced to address endogeneity concerns: percentage of the gross irrigated area ($Irri_{(ds)ry}$) and yields from the previous year ($Yield_{(ds)r(y-1)}$) for weather and pest queries, respectively.

The vector $X_{(ds)ry}$ represents a collection of baseline control variables. These variables are designed to encapsulate a range of characteristics pertaining to the biophysical environment of the districts. This includes elements such as weather conditions, the nature and extent of agricultural land, the utilization patterns of fertilizers, and demographic details concerning the rural population. Meanwhile, the vector $Y_{(ds)ry}$ encompasses a set of additional control variables. These are intended to quantify and describe the level of infrastructure support available for agriculture within the district. Specifically, these controls reflect the presence and quality of infrastructure such as local markets, urban governing bodies, the extent of road networks, and the availability of train stations, all of which facilitate agricultural activities. The fixed effects of the district-State ($\eta_{(ds)}$) and season (ψ_r) are included in the analysis to account for the unobserved heterogeneity at these levels. Standard errors are clustered at the district $\times$ season level to address the potential correlation within the same district and season.

3.2 Discussion on endogeneity issues and first stage regression

Rainfall is one of the critical factors in Indian agriculture, especially for paddy cultivation. Therefore, weather-related queries, especially regarding rainfall predictions, are instrumental in assisting farmers with their cropping decisions. In the era of climate change (i.e., recording erratic rainfall throughout the seasons), the KCC weather prediction services facilitate farmers for paddy crops and schedule irrigation, which play a vital role in meeting crop water needs. Therefore, the percentage of gross irrigated area is correlated with the percentage of seasonal weather queries in each district, making it a suitable instrumental variable to analyze the effect of incidences of weather-related queries on seasonal paddy yields.

The second most frequent query category is related to pest control, indicating the regional infestation of pests (i.e., insects and pathogens that cause economic damage to paddy crops). The incidence of pest-related queries is correlated with seasonal yields in the previous year, as farmers may have experienced challenges in achieving their potential harvests. This experience motivates farmers to seek advice from KCC to minimize yield losses in the current cropping year. Consequently, the crop yields of previous years serve as a suitable instrumental variable for analyzing pest-related queries.

By employing an instrumental variables approach, we address potential endogeneity concerns and lay the groundwork for our first-stage regression analysis. The instruments should exhibit statistically significant relationships with the treatment variables, and the corresponding first-stage F-statistics should exceed the conventional threshold of 10, confirming their relevance. The first-stage regression equations are specified as follows:

$$\begin{aligned} Queries_pest_{(ds)ry} =& \alpha + \beta_1 Yield_{(ds)r(y-1)} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \eta_{(ds)} \times \psi_r + \lambda_y + \epsilon_{(ds)ry} \end{aligned} \tag{2}$$

$$\begin{aligned} Queries_weather_{(ds)ry} =& \alpha + \beta_2 Irri_{(ds)ry} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \eta_{(ds)} \times \psi_r + \lambda_y + \epsilon_{(ds)ry} \end{aligned} \tag{3}$$

The dependent variables in equation (2) and equation (3) represent the percentage of seasonal queries at the district level on pests and weather, respectively. The coefficient β_1 represents the correlation between the incidence of pest-related queries and the yields of the previous year. Similarly, the coefficient β_2 captures the correlation between the percentage of gross irrigated area and the incidence of weather-related queries. The compositions of controls and fixed effects variables are similar to the equation (1), including year fixed effect in each first-stage regression.

4 Identification

The critical identification challenge is the potential endogeneity resulting from the non-random use of KCC services for weather and pest-related issues for paddy. The empirical literature on the nexus between information for agriculture and crop yields assumes that the relationship is unidirectional, that is, the information demand affects crop yield, but not vice versa. For example, Fabregas et al. (2019) concluded that mobile phones can access locally relevant weather and pest information, and their empirical analysis suggests that agricultural information provided through mobile phones increased crop yields by 4% and hence income. Again, Feder and Slade (1984); Suri (2011); Abdul-Rahaman and Abdulai (2020) infers that large and resourceful farmers adopt farming knowledge better than small and resource-deficient farmers. Agricultural extension services play a vital role in adopting technologies in developing countries (Abdulai, 2023). In the context of our analysis, Indian farmers are mostly small and marginal farmers (i.e., farm size less than 2 hectares); one valid concern is the simultaneity bias between the incidence of calls and the yield of the paddy in the equation (1). A recent study on the growth of mobile phone diffusion in India suggests that electricity consumption, Human Development Index (HDI), Herfindahl Index (HHI), and the proportion of rural to urban ratio do not play a significant role in mobile phone penetration in India (Mukhopadhyay et al., 2023). The study also infers that States with more rural regions have higher adoption rates of mobile phone technologies due to lower tariffs. Therefore, the accessibility of mobile phone technologies is not a constraint in accessing digital agricultural information in rural India.

Our analysis involves a longitudinal database observed at the district level for 11 years. However, we can control for fixed effects at the farmers' level characteristics. Our analysis includes tests for spatial and serial correlation and clustering standard errors at the district-State level (Ta-

ble B.2). We also present the results based on Driscoll-Kraay clustered standard errors to address any serial correlation issues as part of robustness checks (Table 8). We include a set of controls as detailed as possible in our setting; nevertheless, there will still be some unobserved characteristics that our controls fail to capture, leading to an identification threat due to omitted variable bias. We address these concerns via the use of two instrumental variables (IVs): (i) percentage gross irrigated area for weather-related queries, as information provided to farmers contributes to the irrigation scheduling decisions, and (ii) previous year's seasonal paddy yields for pest-related queries, as pest management experiences contribute to the current year's paddy yield, reflected in the frequency of queries related to pests from farmers' fields. It is important to present how these choices of these instruments address the issues related to the omitted variable bias and improve the coefficient estimates of the treatments. We present a detailed analysis based on Oster δ omitted variable bias test (Table B.3) and decomposing IV-OLS coefficient gaps (Table B.4). These are discussed below:

4.1 Previous-year's seasonal yields

Pest infestations in crops are the second most important factor after weather that influences Indian agriculture, as recorded through KCC queries from 2009 to 2019 (Table 1). Pest infestations, including pathogens and insects that cause economic damage to crops, are specific to the crop and growth stages in the life cycles of the crops. The pests feed on plant parts and affect the plants' growth, thus reducing the crop yield. Pest infestations evolve over seasons and years for the respective crops due to the evolving nature of pest biology, given the environmental conditions. Farmers' experiences with the yields of previous years affect the current year's farmers' queries at KCC on pests for the respective crops. The KCC, being equipped with advanced knowledge of pest control measures, serves farmers efficiently, as observed in the analysis. The seasonal paddy yields for the previous years are significantly and positively correlated with the percentage of pest-related queries at the district level (Table B.5, Figure B.1, Figure B.2). These first-stage regression results for the full, region and state-wise samples imply that the experience of the increase in yield in previous years' paddy yields due to information support from KCC influenced the increase in current year district-level percentages of pest-related queries. Moreover, the F-statistic values in the first stage-regressions with baseline controls and including both baseline and additional controls are 11.75 and 11.97, respectively, emphasizing the relevance of this instrument (Table B.5).

Thus, the yield of the previous year is selected as the instrumental variable for the pest-related treatment variable.

4.2 Percentage gross irrigated area

Indian agriculture is largely dependent on rainfall. The main staple, paddy, depends on rainfall or irrigation facilities. The paddy crop is mainly grown during the monsoon season in India and depends on rainfall. The rainfall deviations affect the activities of paddy crop management. The alternative water source for cultivation is irrigation. Farmers schedule irrigation and other cropping activities, having knowledge of the arrival of monsoons and the intensity of rainfall. In the era of climate change, erratic rainfall in the winter season affects crop growth and the decision about irrigation scheduling. Thus, an agriculture advice system such as KCC plays a vital role for farmers in India in making informed decisions about their irrigation scheduling and other cropping activities. The incidences of weather-related queries in KCC, which are mostly related to rainfall forecasts, were significantly and negatively correlated with the percentage of gross irrigated area for the full, region, and state-wise samples (Table B.5, Figure B.1, Figure B.2). Farmers who have irrigation systems tend to obtain less weather-related information from KCC, as suggested by these first-stage regressions. The relevance of this chosen instrument is evident as the F-statistic is greater than 10 (32.67 and 33.50) across different specifications (only baseline control and both baseline and additional controls) in the full first-stage regression (Table B.5). Therefore, the variable is chosen as an instrumental variable for the percentage of weather-related queries at the district level.

4.3 Validity of instruments

The analysis involves a series of tests to corroborate the choice of instruments. Before checking if the IV assumptions are satisfied, we would like to test if there are spatial and serial correlations in both outcome and treatment variables, given our analysis is based on longitudinal agricultural data at the district-level. We tested for spatial autocorrelation in the IV regression residuals using Morans I under randomization with a 4-nearest-neighbors weight matrix. The statistic ($I = 0.0715$, $p = 1.00$) indicates no significant spatial dependence, suggesting that the residuals are spatially random after controlling for district-level effects and covariates. Now, if there is a highly signif-

icant serial correlation, then the exclusion restriction is unlikely to hold (Tsay, 2005). We present the state-level autocorrelation function plots for the outcome variable (i.e., paddy yield) and treatment variables (i.e., percentage of farmers' queries at KCC for paddy pests and weather) (Figure 5), and we have found no significant serial correlation since the bar graphs are within the confidence intervals. Next, we formally tested for serial correlation in the panel fixed-effects residuals using the Wooldridge (Drukker, 2003), BaltagiLi LM, and DurbinWatson tests. The Wooldridge test (F = 6.14, p = 0.013) and LM test (χ^2= 4.47, p = 0.034) reject the null of no first-order autocorrelation, while the DurbinWatson variant (DW = 2.08, p = 0.98) does not (Table B.2). Considering the residuals as exhibiting mild AR(1) dependence, we therefore compute DriscollKraay standard errors, which are robust to serial and cross-sectional correlation as well as heteroskedasticity (Table 8). Next, we proceed to the test for instruments. Moreover, we regress the instruments against all the control variables. Should the instruments display a strong correlation with the observables, it is probable they are also correlated with the unobservables (Altonji et al., 2011). The majority of the control variables do not exhibit significant correlations, while some, such as fertilizer consumption, show only weak significant correlations with the gross irrigated area. This affirms that the chosen instruments are appropriate. (Table B.6).

The three identification assumptions of instrumental variables are the endogeneity of the regressors (OLS and 2SLS estimators are consistent when the regressors are exogenous, but if one or more regressors are endogenous, then the OLS estimator is inconsistent), relevance (instruments are correlated with the endogenous regressor) and exclusion restriction. Firstly, we perform the Wu-Hausmann test of endogeneity to verify if the OLS and 2SLS estimators are consistent. A significant result suggests inconsistent OLS and consistent 2SLS results, implies that the 2SLS estimating procedure is relevant, and confirms the presence of endogeneity with one or more regressors. Thus, we ensure that the inclusion of instruments is relevant in the analysis (Thurman, 1986; Hansen, 1982; Altonji et al., 2005). Second, the relevance test of the instrumental variable is evaluated based on the weak instrument statistic (Bound et al., 1995; Staiger and Stock, 1994; Stock and Yogo, 2002; Kleibergen and Paap, 2006). Third, after we test the relevance of instruments and endogeneity of regressors, the exclusion restriction (instruments affect the outcome only through endogenous regressors) is not directly testable. As a result, in subsection 5.2, we present additional evidence that the selected instruments affect paddy yields mainly through treatment variables (i.e., percentage seasonal queries on pests and weather at the district level) in our context.

Lastly, we present some additional tests such as falsification, placebo and non-linear (threshold effects on treatment) tests. For the falsification test, we consider wheat and pulses yields as outcome variables, for the placebo test, we used next-year paddy pest and weather-related queries as treatments, and for the nonlinear tests, we used the Quadratic model and cubic B-splines.

4.4 Monotonicity

When estimating local average treatment effects (LATE) with IV specifications, it is crucial to verify the monotonicity assumption, which posits that individuals influenced by the instruments are impacted in a similar manner (Imbens and Angrist, 1994; Angrist and Pischke, 2009). The monotonicity assumption tests[27] implies that the direction of the relation between the treatment variables and their respective instrumental variables remains true in different sample groups. We present the analysis results under the first-stage regression table Table B.1 and the subgroups based on region (Figure B.1) and States (Figure B.2). The directions of the coefficient estimates are consistent: Previous years' seasonal paddy yields (IV) positively correlated with percentage district-level paddy pest-related queries (treatment), and percentage gross irrigated area negatively correlated with percentage district-level weather-related queries (treatment). Additionally, Figure B.2 illustrates the partial residuals of the treatment variables in relation to the IVs, detailing how a one-unit change in the IVs affects the corresponding change in the treatment variables both for individual States and when all States are considered together. These plots affirm that our IV strategy meets the monotonicity assumption.

To assess the validity of the monotonicity assumption underlying our instrumental variable (IV) strategy, we examine the sign and distribution of the 2SLS weights associated with each instrument. Monotonicity requires that the instruments influence the treatment in a consistent direction across all compliersi.e., the weights should be non-negative (or non-positive) depending on the instruments effect. This condition is central to the identification of Local Average Treatment Effects (LATE) in IV models, as formalized by Imbens and Angrist (1994) and further elaborated by Angrist and Pischke (2009). Table B.7 presents the results of this test for both continuous and binary specifications of the treatment and instrument variables.

[27]The monotonicity assumption intuitively suggests that individuals influenced by the instruments should experience comparable effects. In technical notations, this is expressed as $D_q(Z_q = 1) \geq D_q(Z_q = 0)$ or $D_q(Z_q = 1) \leq D_q(Z_q = 0)$ for any KCC query type (q), where D_q represents the endogenous regressor and Z_q indicates the instrument vector (Angrist and Pischke, 2009).

In the continuous specification, the instrument for pest-related queriesprevious years seasonal paddy yieldhas a positive first-stage coefficient (1.041, significant at the 10% level). However, the associated 2SLS weights are nearly evenly split, with 48.7% positive and 51.3% negative, suggesting that the monotonicity condition may not be strictly satisfied in this case. In contrast, the instrument for weather-related queriespercentage of gross irrigated areahas a negative and significant first-stage coefficient (-0.020), and all associated 2SLS weights are positive ($p = 1.000$), with none negative ($p = 0.000$), supporting the monotonicity assumption.

To further probe robustness, we re-estimated the model using binary versions of both the treatment and instrument variables, defined as indicators for values above the sample mean. In this binary specification, the results are more supportive of monotonicity. The instrument for pest-related queries (binary yield) has a highly significant positive effect (0.079, $p < 0.01$), and 99.5% of the 2SLS weights are positive. Similarly, the instrument for weather-related queries (binary irrigation) yields a negative coefficient (-0.320) with 100% of the weights positive and none negative. These results strongly support the monotonicity condition in the binary framework.

Together, these findings suggest that while monotonicity may be imperfect in the continuous specification for pest-related queries, it is well-supported in the binary specification and for weather-related queries across both models. This reinforces the credibility of our LATE interpretation and the robustness of our IV strategy.

4.5 Causal pathway and agronomic logic for instrumental variable strategy

To enhance the interpretability of our identification strategy, we include a conceptual path diagram (Figure 8) that visually maps the causal relationships underlying our instrumental variable approach. The diagram shows how the instrumentsprevious seasons paddy yield and percentage of gross irrigated areaaffect farmers decisions to seek pest- and weather-related advice through the KCC. These treatment variables, in turn, influence current paddy yields. The diagram also highlights the role of unobserved confounders that may simultaneously affect query behavior and yield outcomes. By leveraging instruments that are plausibly exogenous to current yield shocks, our IV strategy aims to recover causal estimates of the effect of information-seeking behavior on agricultural productivity. This visual framework helps clarify the logic of our empirical design and its relevance for policy audiences interested in the role of advisory services in agriculture.

To complement the econometric validity of our IV strategy, we also consider the agronomic rationale linking the instruments to farmers information-seeking behavior. The percentage of gross irrigated area serves as a proxy for a districts resilience to weather variability. In areas with limited irrigation coverage, farmers are more exposed to rainfall uncertainty and extreme weather events, which increases their likelihood of seeking weather-related advice through platforms like the KCC. Conversely, districts with higher irrigation coverage may experience fewer disruptions, reducing the need for such queries. Similarly, previous seasons paddy yield reflects recent agronomic performance and pest pressure. A poor yield may signal unresolved pest issues or heightened vulnerability, prompting farmers to seek pest-related guidance in the subsequent season. These instruments thus capture structural agronomic conditions that plausibly influence farmers decisions to engage with advisory services, reinforcing the relevance of our identification strategy.

5 Results

We present the effect of the seasonal percentage of district-level queries of farmers related to paddy pests and weather at KCC on seasonal paddy yields in the following subsections. These query types are not significantly correlated to each other (Correlation coefficient: -0.015, p-value: 0.157).

5.1 The Impact of query types on seasonal crop yields

We first present the results of the ordinary least squares (OLS) regressions and then proceed to the main regression results using the 2SLS/IV estimation, including the first-stage regression.

The OLS and 2SLS/IV regression results are reported in the third and sixth columns of Table 6, respectively. The other columns (first, second, fourth, and fifth) present estimates when the treatment variables (pest and weather) are individually regressed with crop yields and are tabulated for comparison with the main regression results. The OLS regression findings indicate that a 1% increase in seasonal district-level queries related to paddy pests increases paddy yield by 0.002 tonnes per hectare (standard error, SE: 0.001). On the other hand, a percentage increase in seasonal weather-related questions at the district level for paddy improves the yield of paddy by 0.007 tonne per hectare (SE: 0.001).

The 2SLS/IV regression results reveal that the Wu-Hausman tests are statistically significant,

indicating the presence of endogeneity in the model. This difference between the OLS and 2SLS/IV estimates highlights the need to address the endogeneity issue. The tests for weak instruments also yield significant results for all specifications, allowing us to reject the null hypothesis and affirm that our instruments are valid. The first-stage regression estimates for paddy, as shown in equations (2) and (3), demonstrate a significant correlation between the treatment variables and the instrumental variables (see the tables in the appendix, subsection B, (Table B.5) and the F-satistic estimates are greater than 10 (Pest-related Queries vs. Previous years' yields: 11.97 and Weather-related queries vs. Percentage gross irrigated area: 33.50).

The absolute values of the 2SLS/IV estimates are larger than those from the OLS regressions. This divergence may reflect attenuation bias in the OLS estimates due to classical measurement error in the treatment variables, which IV estimation is designed to correct under standard assumptions (Angrist and Krueger, 2001). While we cannot directly test for measurement error in this framework, the IV estimates likely provide more reliable causal effects in the presence of such bias. This reduction is also influenced by the variability of the treatment variables relative to their respective instruments. Specifically, for IV estimates, an increase in the unit percentage of seasonal queries related to pests at the district level due to pest infestations in paddy fields leads to a decrease in paddy yield of 0.028 t ha^{-1} (SE: 0.014). On the other hand, an increase in the unit percentage of seasonal queries related to weather at the district level results in an improvement in paddy yield of 0.008 t ha^{-1} (SE: 0.005). The increase in distress queries related to pests decreases 1.1% of the average seasonal paddy yields (i.e., 2.6 t ha^{-1}, Table 3), while an increase in weather-related queries increases average seasonal paddy yields by 0.3%. While this effect may appear modest in absolute terms, it is economically and agronomically meaningful when scaled across large cultivation areas. Even small improvements in yield per hectare can translate into substantial aggregate gains in production and income, particularly in regions with high paddy acreage. Moreover, the positive effect of weather-related queries may reflect improved farmer preparedness and timely decision-making in response to climate variability, underscoring the practical value of accessible agro-advisory services.

To explore the reasons for the differences between OLS and IV estimates, we apply two diagnostic approaches: the Oster δ-robustness test for omitted variable bias and the IV-OLS gap decomposition method proposed by Ishimaru (2024). The Oster test evaluates how much selection on unobservables would be required to reconcile the OLS estimate with the IV estimate (Table

B.3). For pest-related queries, the δ value is 1.56, indicating that unobservables would need to be more influential than observables to explain the gap – suggesting that the OLS estimate is relatively robust and the IV estimate reflects a stronger causal effect. In contrast, the δ value for weather-related queries is 0.27, implying that the OLS estimate is more sensitive to omitted variables and that the IV estimate is likely more reliable. This test follows the framework developed by Oster (2019), which links coefficient stability to selection bias.

To further unpack these differences, we decompose the IV-OLS gap (Table B.4) into three components following the empirical strategy of Ishimaru (2024): differences in covariate weights (ΔCW), treatment weights (ΔTW), and marginal effects (ΔME). For pest-related queries, the IV estimate is substantially more negative than the OLS estimate, with most of the gap driven by the marginal-effect component (ΔME = 0.0290). This suggests that the treatment effect among compliersthose whose treatment status is influenced by the instrumentis more negative than the average effect captured by OLS. For weather-related queries, the IV and OLS estimates are directionally consistent, and the small gap is primarily explained by differences in marginal effects (ΔME = 0.0075), while treatment-weight differences (ΔTW = 0.0063) indicate that compliers differ meaningfully in their exposure to the treatment. These findings reinforce the importance of IV estimation in uncovering heterogeneous treatment effects and highlight the limitations of relying solely on OLS when unobserved confounding or treatment heterogeneity is present.

The negative association between pest-related queries and paddy yields may reflect underlying reporting bias in the use of the KCC platform. Specifically, farmers are more likely to seek assistance when facing acute pest infestations or crop distress, which could lead to a concentration of queries among those experiencing adverse agronomic conditions. In this context, query volume may not represent general pest prevalence or proactive information-seeking behavior, but rather a reactive signal of localized crop damage. This behavioral pattern introduces a form of selection bias, where the treatment variable disproportionately captures distress-driven demand. While our instrumental variable strategy helps mitigate endogeneity concerns, we acknowledge that the interpretation of pest-related query volume may be confounded by selective reporting.

5.2 Assessing the exclusion restriction

In this subsection, we assess the validity of the exclusion restriction of the chosen instruments through the following tests.

5.2.1 Zero-first-stage test

First, we performed the zero first stage test (an informative test) to establish that the exclusion restriction is satisfied (Van Kippersluis and Rietveld, 2018). This methodology was originally presented by (Angrist and Krueger, 1995) and subsequently applied by (Bound and Jaeger, 2000); it was later referred to by (Altonji et al., 2005; Angrist et al., 2010). The intuition behind this approach is that the instrument should have no direct effect on the outcome in a sub-sample where it has no impact on the treatment variable if the exclusion restriction holds. Although this analysis cannot definitively validate the exclusion restriction, it can help assess the plausibility of the assumption. We selected the State sub-samples based on the frequencies of particular types of farmers' queries, cropping patterns, and utilization of irrigation facilities.

Farmers' queries from the eastern (Bihar and Odisha) and western (Gujarat and Maharashtra) regions are predominantly focused on schemes, financial support, and contact information to access government benefits. Together, these query types account for approximately 28% of the total queries, ranking second only to weather-related queries in the respective regions (Figure 1). Besides the nature of queries, in the western region, such as Gujarat farmers focus less on paddy cultivation (see seasonal paddy cultivation maps and discussion in the appendix, subsection C) and Maharashtra lacks irrigation support (Figure C.4). The incidence of queries related to pests and weather does not significantly affect paddy yields in the districts of these regions (i.e., we have found an insignificant correlation between instrument variables, treatment variables, and outcome variable). On the other hand, the districts of Punjab and Haryana have already achieved their yield potential (i.e., maximum possible output from its given sets of inputs including genetic potential of the plants, following the definition of technical efficiency); thus, we don't expect a significant yield increase due to information access from KCC or other interventions. The zero first-stage samples are selected on the basis of the individual instrument variables. The districts of Bihar, Odisha, Gujarat, Haryana, and Punjab (consisting of $\sim 47\%$ of observations) are selected for the previous years' paddy yields, which is the IV for pest-related queries. For the percentage

of the gross irrigated area (i.e., IV for weather-related queries), districts of Gujarat Maharashtra and Punjab are selected (accounting for $\sim$ 30% of observations).

Table B.8 present the results of this analysis. The regression results are presented in columns (1) and (2) for the treatment variables and queries at the KCC related to pests and weather. Columns (3) and (4) present the results with the outcome variable (i.e., current year yield) with the two instrumental variables (i.e., the yield from the previous year for pest-related queries and the percentage of gross irrigated area for weather-related queries). The results for the full sample, the zero-first-stage sample group, and the remaining sample group (i.e., full sample excluding the zero-first-stage sample group) show that the effects of the instruments on treatment and outcome variables are statistically significant under full and remaining samples but statistically insignificant under the zero-first-stage sample groups. However, the correlation between our chosen instrumental variable for pest-related queries (previous years' seasonal paddy yields) and the outcome variable (current years' seasonal yields) is not statistically significant, which is consistent with our serial correlation plots based on States (Figure 5). It should be noted that the F-statistics for all regressions are greater than 10, including those under the zero-first-stage sample groups.

5.2.2 Additional indirect tests:

Secondly, we test the determinants of crop yields, other than the incidences of query types at KCC, through the following sets of equations.

$$\begin{aligned} \textit{NPK per GCA}_{(ds)y} =& \alpha + \beta_1 Yield_{(ds)r(y-1)} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \lambda_y + \eta_{(ds)} \times \psi_r \times \lambda_y + \epsilon_{(ds)ry} \end{aligned} \tag{4}$$

$$\begin{aligned} Rainfall_{(ds)ry} =& \alpha + \beta_1 Irri_{(ds)ry} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \lambda_y + \eta_{(ds)} \times \psi_r \times \lambda_y + \epsilon_{(ds)ry} \end{aligned} \tag{5}$$

$$
\begin{aligned}
Paddy\ Price\ Shock_{(ds)y} =& \alpha + \beta_1 Yield_{(ds)r(y-1)} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\
&+ \eta_{(ds)} + \psi_r + \lambda_y + \eta_{(ds)} \times \psi_r \times \lambda_y + \epsilon_{(ds)ry}
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
Paddy\ Price\ Shock_{(ds)y} =& \alpha + \beta_1 Irri_{(ds)ry} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\
&+ \eta_{(ds)} + \psi_r + \lambda_y + \eta_{(ds)} \times \psi_r \times \lambda_y + \epsilon_{(ds)ry}
\end{aligned} \tag{7}
$$

The total fertilizer consumption (NPK) per GCA is regressed with the pest-related instrumental variable, the previous year's yield, in equation (4). The weather-related instrumental variable, the percentage gross-irrigated area, is regressed with the rainfall in equation (5), and both the instrumental variables are regressed with the paddy price shocks in equations (6) and (7), respectively. The composition of control variables and fixed effects remains similar as described in equation (2).

Table B.9 reports the results of the OLS estimation of equations (4), (5), (6) and (7). It is clear that our chosen instruments are uncorrelated with variables commonly expected to be associated with paddy yield, i.e., total fertilizer consumption per gross cropped area and rainfall. The baseline results under Table B.10 (column 3) reveal significant correlations with the outcome variable (i.e., current year yield) in the main regression equation (1) the selected outcome variable rainfall in equation (5). The other two factors, namely the percentage of gross irrigated area and the total fertilizer use per gross cropped area, exhibit minor insignificance concerning the paddy yields for the current year. These findings support the exclusion restriction in our context. The selected instruments do not imply significant variations in the determinants of paddy yields other than through the incidence of queries through KCC.

5.3 Results on additional tests: Falsification, Placebo and Non-linearity

To assess the validity and robustness of our main findings, we conduct a series of falsification, placebo, and nonlinearity checks, summarized in Table B.11. The falsification tests use alternate dependent variableswheat and pulses yieldsto examine whether pest and weather-related queries have unintended associations outside the paddy context. The results show no statistically significant effects, supporting the specificity of the treatment-outcome relationship for paddy. Placebo

tests using future-season queries (i.e., pest and weather queries from season yield null effects on current paddy yields, reinforcing the temporal validity of our identification strategy.

We also explore potential nonlinearities in the relationship between query incidence and paddy yield using cubic B-splines and quadratic specifications. The spline-based estimates reveal no consistent pattern across low, medium, and high query intensities, and the quadratic terms are statistically insignificant. These results suggest that the effects of query behavior on yield are approximately linear within the observed range. Overall, the robustness checks affirm that the observed treatment effects are not driven by spurious correlations, reverse causality, or model misspecification.

6 Robustness checks

Table 7 presents a series of robustness checks to validate the consistency of our main findings. First, we re-estimate the OLS and IV models using Driscoll-Kraay standard errors to account for potential spatial and temporal dependence in the error structure. The results remain consistent with our IV estimates in Table 6: pest-related queries continue to show a negative and statistically significant effect on paddy yields in the IV specification (0.029, SE: 0.017), while weather-related queries are associated with a positive yield effect (0.007 to 0.008, SE: 0.006). Second, we augment the model with controls for district-level price shocks to account for market-driven incentives, pest outbreaks, and weather shocks that may influence both query behavior and yield outcomes. The inclusion of these controls does not materially alter the magnitude or significance of the coefficients. Across all specifications, the direction and statistical significance of the treatment effects remain stable, reinforcing the robustness of our identification strategy and the reliability of the estimated impacts.

Next, Table 8 presents robustness checks that incorporate a range of extended fixed effects and control variables to test the stability of our IV estimates. Across all 11 model specifications, the negative effect of pest-related queries on paddy yields remains consistently significant, with coefficients ranging from 0.029 to 0.033. Similarly, the positive effect of weather-related queries is stable across specifications. These models sequentially introduce agro-climatic zone[28] fixed

[28]The districts are mapped to eleven agro-climtaic zones: Central Plateau and Hills, Eastern Coastal Plains and Hills, Eastern Plateau and Hills, Gujarat Plains and Hills, Middle Gangetic Plains Region, Southern Plateau and Hills, Trans-

effects, landholding type[29], mobile boom periods, and a rich set of controls including weather z-scores, mobile ownership by gender, nightlight intensity, land ownership, and literacy rates. The Wu-Hausman statistics remain highly significant throughout, confirming the relevance of the IV strategy. Overall, these results demonstrate that the estimated effects are robust to a wide array of agronomic, demographic, and infrastructural controls, reinforcing the credibility of our identification approach.

Table 9 presents additional robustness checks to test the sensitivity of our IV estimates to alternate specifications. Columns (1) and (2) uses relative grain yield (RGY)[30] for paddy and log-transformed paddy yield as the dependent variable. yielding consistent signs but statistically insignificant effects. In column (2) we use alternate IVs (i.e., 1. Average of two-year lagged paddy yield as IV for pest-related queries, and 2. Five-year rolling CAGR of percentage gross irrigated area as IV for weather related queries), showing similar coefficient estimates for pest and weather-related queries -0.006 (SE: 0.010) and 0.012 (SE: 0.003), respectively. Next in Column (3), we introduce additional IVs for weather-related queries (i.e., percentage gross irrigated area and five-year rolling CAGR of percentage gross irrigated area), whereas the IV for pest-related queries remain same to our original setting (i.e., previous years' paddy yields).The results indicate a statistically significant negative effect of pest-related queries (0.010, SE: 0.006) and a positive effect of weather-related queries (0.012, SE: 0.003). The Hansen J statistic in column (3) indicates no overidentification concerns (p = 0.90). Columns (4) through (6) estimate seasonal subsamplesmonsoon, winter, and summerrevealing that the negative effect of pest-related queries is strongest and significant in the winter season (0.019 t ha^{-1}), while weather-related queries show no consistent seasonal pattern. Wu-Hausman statistics across specifications support the relevance of the IV strategy. Overall, these checks confirm that the estimated effects are robust to alternative instruments and seasonal heterogeneity.

Ganga Plains Region, Upper Gangetic Plain Region, Western Coastal Plains and Ghats, Western Dry Region, Western Plateau and Hills.

[29]Landholding size groups: small and marginal holdings (0-2 hectares), medium holdings (2-10 hectares), and large holdings (>10 hectares).

[30]Relative Grain Yield (RGY) represents a standardized version of the yield metric (Pearce et al., 2022; Rogers et al., 2025). For our analysis it is computed as the ratio between the seasonal paddy yields in a district of a State and a year, $Yield_{(ds)ry}$, and the maximum seasonal paddy yield recorded across all districts in a State for the corresponding year, $\max(Yield_{(s)ry})$, such that

$$RGY = \frac{Yield_{(ds)ry}}{\max(Yield_{(s)ry})}.$$

This normalization facilitates direct comparison of yield performance across different environments.

7 Conclusion

Extension services facilitated by mobile technologies have shown promise in improving agricultural outcomes in India and other developing countries (Cole and Fernando, 2021; Fabregas et al., 2019). The *Kisan* (farmers) call center (KCC) offers a scalable platform for farmers to access timely agronomic advice via mobile connectivity. Drawing farmers' query level data from KCC and district-level paddy production records across major agricultural states, our analysis identifies a robust and conditionally causal relationship between seasonal paddy yields and the incidence of pest- and weather-related farmer queries.

We find that higher volumes of pest-related queries are associated with lower paddy yields, likely reflecting reactive information-seeking behavior in response to crop damage. In contrast, weather-related queries are positively associated with yield, suggesting that proactive engagement with climate information may support better preparedness and decision-making. These differential effects underscore the importance of tailoring advisory content to the timing and nature of farmer needs.

While our instrumental variable strategy helps address endogeneity concerns, we acknowledge limitations in fully ruling out residual confounding and reporting bias. The interpretation of query volume as a proxy for farmer behavior remains conditional on the validity of the instruments and the structure of the advisory system.

Policy implications extend beyond generic calls to strengthen KCC. We recommend targeted enhancements such as AI-based triage of incoming queries, region-specific pest alerts, and integration with localized weather forecasting. A detailed cost-benefit analysis comparing digital advisory platforms like KCC with traditional extension models would further inform resource allocation. But mobile phone-based information access through KCC and subscription to the KCC services definitely reduce several costs (i.e., search, replication, transportation, tracking, and verification costs as identified as a benefit of using digital economic tools by Goldfarb and Tucker (2019)) as compared to the traditional in-person extension services. Lessons from SMS-based advisory programssuch as Kenyas iShamba service and randomized trials in Kenya, Nigeria, Nepal, and Indiademonstrate the value of personalized, timely, and low-cost messaging for farmer decision-making (Tabe-Ojong Jr et al., 2023; Casaburi et al., 2014; Giulivi et al., 2023; Fabregas et al., 2019).

Future research should explore the effectiveness of KCC across other staple and cash crops, while also conducting a detailed study of gendered patterns of mobile advisory adoption and the implications of the digital divide, to enrich policy prescriptions.

References

Abdul-Rahaman, A. and Abdulai, A. (2020). Social networks, rice value chain participation and market performance of smallholder farmers in ghana. *African Development Review*, 32(2):216–227.

Abdulai, A. (2023). Information acquisition and the adoption of improved crop varieties. *American Journal of Agricultural Economics*, 105(4):1049–1062.

Afridi, F., Bishnu, M., and Mahajan, K. (2023). Gender and mechanization: Evidence from indian agriculture. *American Journal of Agricultural Economics*, 105(1):52–75.

Agrawal, S., Harish, S., Mahajan, A., Thomas, D., and Urpelainen, J. (2020). Influence of improved supply on household electricity consumption-evidence from rural india. *Energy*, 211:118544.

Aker, J. C. (2010). Information from markets near and far: Mobile phones and agricultural markets in Niger. *American Economic Journal: Applied Economics*, 2(3):46–59. Publisher: American Economic Association.

Aker, J. C. (2011). Dial a for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agricultural economics*, 42(6):631–647.

Aker, J. C., Ksoll, C., and Lybbert, T. J. (2012). Can mobile phones improve learning? evidence from a field experiment in niger. *American Economic Journal: Applied Economics*, 4(4):94–120.

Altonji, J. G., Conley, T., Elder, T. E., and Taber, C. R. (2011). Methods for using selection on observed variables to address selection on unobserved variables (preliminary and incomplete).

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human resources*, 40(4):791–821. Publisher: University of Wisconsin Press.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple Experiments for the Causal Link between the Quantity and Quality of Children. *Journal of Labor Economics*, 28(4):773–824.

Angrist, J. D. and Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, 13(2):225–235.

Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. In *Econometric Society Monographs*, volume 33, pages 1–81. Cambridge University Press.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Antweiler, W. and Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3):1259–1294.

Bandumula, N. (2018). Rice production in asia: Key to global food security. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, 88(4):1323–1328.

Binswanger, H. P., Khandker, S. R., and Rosenzweig, M. R. (1993). How infrastructure and financial institutions affect agricultural output and investment in India. *Journal of development Economics*, 41(2):337–366. Publisher: Elsevier.

Blumenstock, J. E., Eagle, N., and Fafchamps, M. (2016). Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157–181.

Bound, J. and Jaeger, D. A. (2000). Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? In *Research in labor economics*, pages 83–108. Emerald Group Publishing Limited.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430):443–450.

Casaburi, L., Kremer, M., Mullainathan, S., and Ramrattan, R. (2014). Harnessing ict to increase agricultural production: Evidence from kenya. *Harvard University*.

Chaurey, A., Ranganathan, M., and Mohanty, P. (2004). Electricity access for geographically disadvantaged rural communitiestechnology and policy insights. *Energy policy*, 32(15):1693–1705.

Cole, S. A. and Fernando, A. N. (2012). The value of advice: Evidence from mobile phone-based agricultural extension. Technical Report Working Paper No. 13-047, Harvard Business School. Working paper.

Cole, S. A. and Fernando, A. N. (2021). Mobileizing agricultural advice technology adoption diffusion and sustainability. *The Economic Journal*, 131(633):192–219. Publisher: Oxford University Press.

Das, S. R. (2012). Rice in Odisha. Technical report, International Rice Research Institute. Available at: https://ageconsearch.umn.edu/record/287645/files/TB16_content.pdf.

Davis, K. and Sulaiman V, R. (2014). The new extensionist: Roles and capacities to strengthen extension and advisory services. *Journal of International Agricultural and Extension Education*, 21(3):6–18.

Deichmann, U., Goyal, A., and Mishra, D. (2016). Will digital technologies transform agriculture in developing countries? *Agricultural Economics*, 47(S1):21–33.

Dominic, D. M., Mitta, N., Sulaiman V, R., Bharti, P., and Sharma, S. (2022). Enhancing capacities of digital extension and advisory services in odisha, india. *CGIAR Research Initiative on Digital Innovation*.

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., and Taneja, J. (2021). Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5):922.

Emerick, K. (2018). Agricultural productivity and the sectoral reallocation of labor in rural india. *Journal of Development Economics*, 135:488–503.

Emerick, K. and Dar, M. H. (2021). Farmer field days and demonstrator selection for increasing technology adoption. *Review of Economics and Statistics*, 103(4):680–693.

Etoori, P., Chinnakotla, M., and Mamidi, R. (2018). Automatic spelling correction for resource-scarce languages using deep learning. In *Proceedings of ACL 2018, Student Research Workshop*, pages 146–152.

Fabregas, R., Harigaya, T., Kremer, M., and Ramrattan, R. (2022). Digital agricultural extension for development. In *Introduction to Development Engineering: A Framework with Applications from the Field*, pages 187–219. Springer International Publishing Cham.

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.

Fan, S., Hazell, P., and Thorat, S. (2000). Government Spending, Growth and Poverty in Rural India. *American Journal of Agricultural Economics*, 82(4):1038–1051.

Feder, G. and Slade, R. (1984). The Acquisition of Information and the Adoption of New Technology. *American Journal of Agricultural Economics*, 66(3):312–320.

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.

Gandhi, V. P. and Johnson, N. (2018). Decision-oriented information systems for farmers: A study of kisan call centres (kcc), kisan knowledge management system (kkms), farmers portal, and m-kisan portal in gujarat. Technical report, Centre for Management in Agriculture (CMA), Indian Institute of Management Ahmedabad, Ahmedabad, India. Supported by the Ministry of Agriculture and Farmers Welfare, Government of India.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, 57(3):535–574.

Giulivi, N., Harou, A. P., Gautam, S., and Guereña, D. (2023). Getting the message out: information and communication technologies and agricultural extension. *American Journal of Agricultural Economics*, 105(3):1011–1045.

Goldfarb, A. and Tucker, C. (2019). Digital economics. *Journal of economic literature*, 57(1):3–43.

Goyal, A. (2010). Information, direct access to farmers, and rural market performance in central India. *American Economic Journal: Applied Economics*, 2(3):22–45. Publisher: American Economic Association.

Gulati, A., Roy, R., and Hussain, S. (2021). Performance of agriculture in gujarat. In *Revitalizing Indian Agriculture and Boosting Farmer Incomes*, pages 113–144. Springer Singapore.

Gupta, A., Ponticelli, J., and Tesei, A. (2024). Language barriers, technology adoption and productivity: Evidence from agriculture in india. *Review of Economics and Statistics*, pages 1–28.

Haile, M. G., Kalkuhl, M., and Von Braun, J. (2016). Worldwide Acreage and Yield Response to International Price Change and Volatility: A Dynamic Panel Data Analysis for Wheat, Rice, Corn, and Soybeans. *American Journal of Agricultural Economics*, 98(1):172–190.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054. Publisher: JSTOR.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Ishimaru, S. (2024). Empirical decomposition of the iv-ols gap with heterogeneous and nonlinear effects. *Review of Economics and Statistics*, 106(2):505–520.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The quarterly journal of economics*, 122(3):879–924. Publisher: MIT Press.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, 133(1):97–126. Publisher: Elsevier.

Komarek, A. M. and Msangi, S. (2019). Effect of changes in population density and crop productivity on farm households in malawi. *Agricultural Economics*, 50(5):615–628.

Kondylis, F., Mueller, V., and Zhu, J. (2017). Seeing is believing? Evidence from an extension network experiment. *Journal of Development Economics*, 125:1–20. Publisher: Elsevier.

Krishna, K. and Meenakshi, J. (2022). Agricultural productivity growth and structural transformation in rural india: Some recent evidence. *Journal of Quantitative Economics*, 20(Suppl 1):277–302.

Kumar, A., Tripathi, G., and Joshi, P. K. (2021). Adoption and impact of modern varieties of paddy in india: evidence from a nationally representative field survey. *Journal of Agribusiness in Developing and Emerging Economies*, 11(3):255–279.

Letta, M., Montalbano, P., and Pierre, G. (2022). Weather shocks, traders' expectations, and food prices. *American Journal of Agricultural Economics*, 104(3):1100–1119.

Li, Y., Miao, R., and Khanna, M. (2019). Effects of ethanol plant proximity and crop prices on land-use change in the united states. *American Journal of Agricultural Economics*, 101(2):467–491.

Lybbert, T. J., Shenoy, A., Bourdier, T., and Kieran, C. (2023). Striving to revive pulses in India with extension, input subsidies, and output price supports . *American Journal of Agricultural Economics*, page ajae.12435.

Miao, R., Khanna, M., and Huang, H. (2016). Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, 98(1):191–211.

Mukhopadhyay, A., Bagchi, K. K., and Udo, G. J. (2023). Exploring the Main Factors Affecting Mobile Phone Growth Rates in Indian States. *Journal of the Knowledge Economy*.

Naldi, M. (2019). A review of sentiment computation methods with R packages. arXiv:1901.08319 [cs].

Nayak, H. S., McDonald, A. J., Kumar, V., Craufurd, P., Dubey, S. K., Nayak, A. K., Parihar, C. M., Peramaiyan, P., Poonia, S., Tesfaye, K., et al. (2024). Context-dependent agricultural intensification pathways to increase rice production in india. *Nature Communications*, 15(1):8403.

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of agricultural science*, 144(1):31–43.

Office of the Registrar General & Census Commissioner, I. (2020). Population projections for india and states, 2011-2036.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. arXiv:1107.4557 [cs].

Parker, C., Ramdas, K., and Savva, N. (2016). Is IT Enough? Evidence from a Natural Experiment in Indias Agriculture Markets. *Management Science*, 62(9):2481–2503.

Pearce, A. W., Slaton, N. A., Lyons, S. E., Bolster, C. H., Bruulsema, T. W., Grove, J. H., Jones, J. D., McGrath, J. M., Miguez, F. E., Nelson, N. O., et al. (2022). Defining relative yield for soil test correlation and calibration trials in the fertilizer recommendation support tool. *Soil Science Society of America Journal*, 86(5):1338–1353.

Rogers, C. W., Pramanik, C., Hu, G., Marshall, J. M., Tarkalson, D. D., Hatzenbuehler, P. L., and Evans, C. (2025). Irrigated malt barley nitrogen management: Insights from historical and modern cultivars. *Soil Science Society of America Journal*, 89(5):e70124.

Scott, S. L. and Varian, H. R. (2015). Bayesian variable selection for nowcasting economic time series. In *Economic analysis of the digital economy*, pages 119–135. University of Chicago Press.

Sharma, S., Kooner, R., and Arora, R. (2017). Insect pests and crop losses. *Breeding insect resistant crops for sustainable agriculture*, pages 45–66.

She, B., Li, H., and Bao, S. (2017). The integration of Internet data and census data for spatial analysis in a geoportal. *Big Data for Regional Science*, page 153. Publisher: Routledge.

Shively, G. and Thapa, G. (2017). Markets, transportation infrastructure, and food prices in nepal. *American Journal of Agricultural Economics*, 99(3):660–682.

Staiger, D. O. and Stock, J. H. (1994). Instrumental variables regression with weak instruments. Technical report, National Bureau of Economic Research Cambridge, Mass., USA.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear IV regression. Technical report, National Bureau of Economic Research Cambridge, Mass., USA.

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, 79(1):159–209.

Tabe-Ojong Jr, M. P., Abate, G. T., Abay, K. A., and Spielman, D. J. (2023). Digital innovations and agricultural transformation in africa: Lessons from kenya.

Thurman, W. N. (1986). Endogeneity testing in a supply and demand framework. *The Review of Economics and Statistics*, pages 638–646. Publisher: JSTOR.

Toyama, K. (2011). Technology as amplifier in international development. In *Proceedings of the 2011 iConference*, pages 75–82.

Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

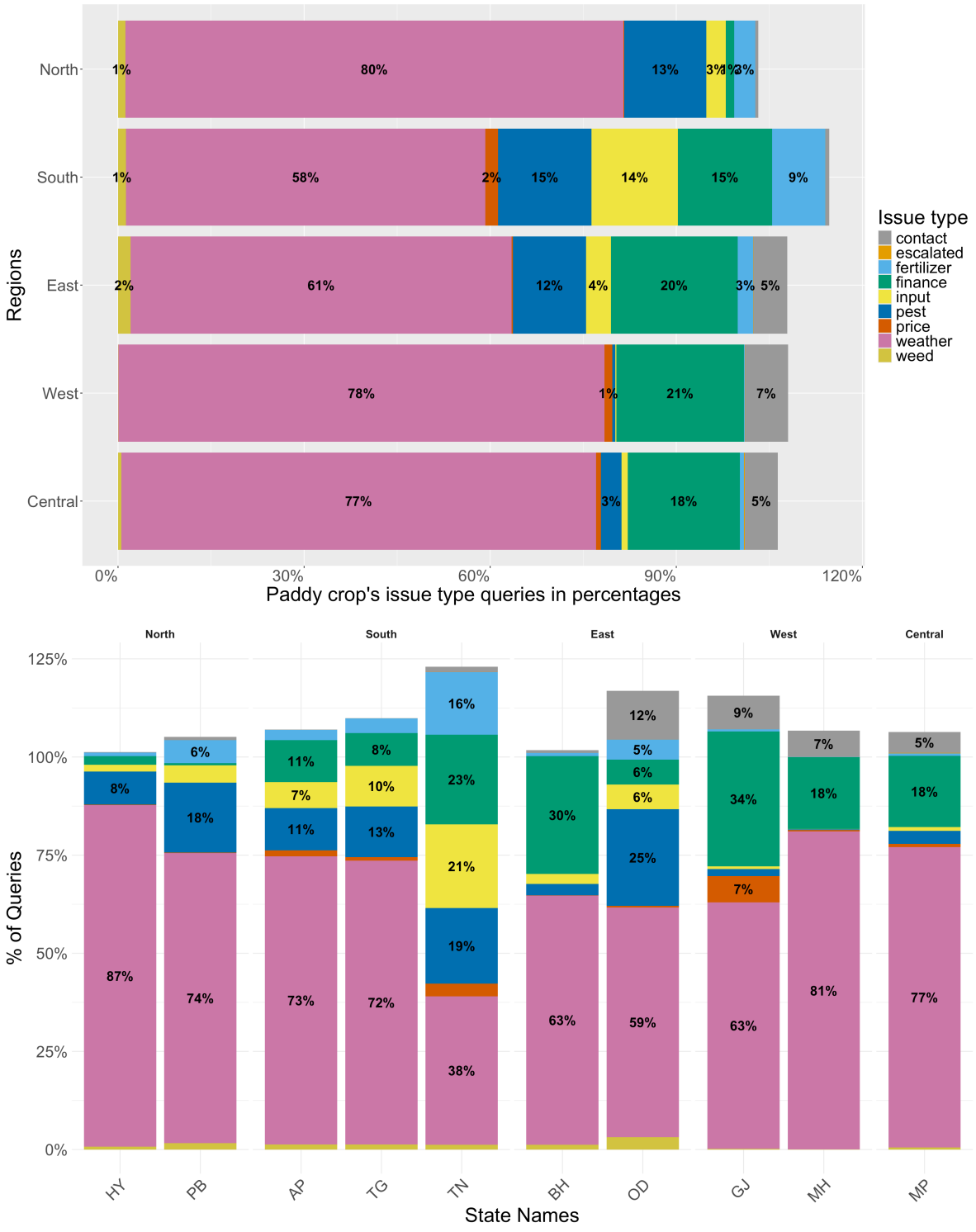
United Nations Population Fund, India (2009). District level population projections in eight selected states of india. Available at: `https://india.unfpa.org/sites/default/files/pub-pdf/unfpa_district_level_population_projections_book_-_with_cover.pdf`.

Van Kippersluis, H. and Rietveld, C. A. (2018). Beyond plausibly exogenous. *The Econometrics Journal*, 21(3):316–331.

Yashavanth, B. S., Sontakki, B. S., Soam, S. K., Venkateshwarlu, G., and Rao, C. S. (2022). Strengthening kisan call centres in india. Technical report, ICAR-National Academy of Agricultural Research Management, Hyderabad, India.

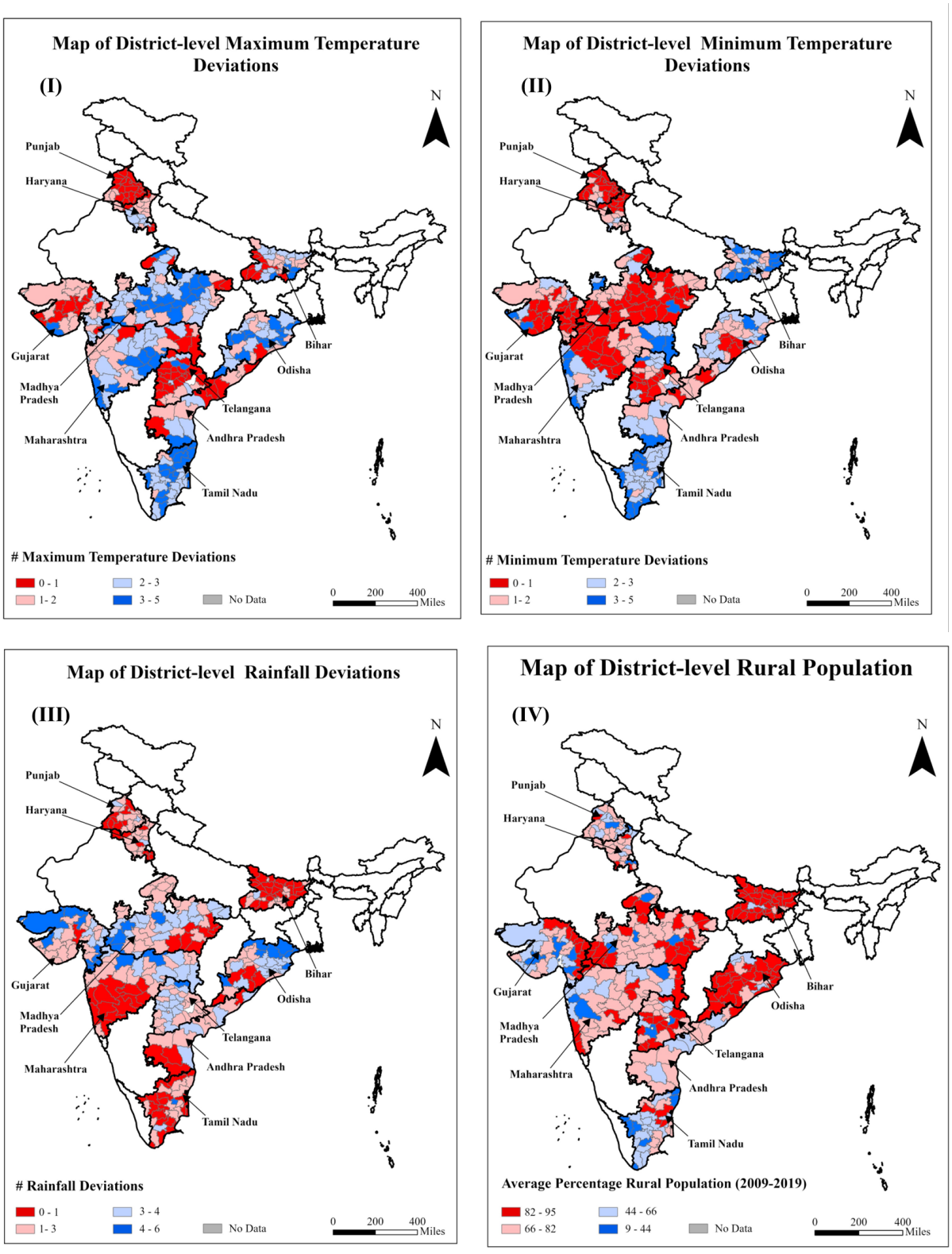
Zhang, X. and Fan, S. (2004). How productive is infrastructure? a new approach and evidence from rural india. *American Journal of Agricultural Economics*, 86(2):492–501.

Figures and Tables



Note: Some of the farmers' queries are on multiple issues and crop categories. Hence, the bars do not always sum to 100%. HY: Haryana, PB: Punjab, AP: Andhra Pradesh, TG: Telangana, TN: Tamil Nadu, BH: Bihar, OD: Odisha, GJ: Gujarat, MH: Maharashtra, MP: Madhya Pradesh.

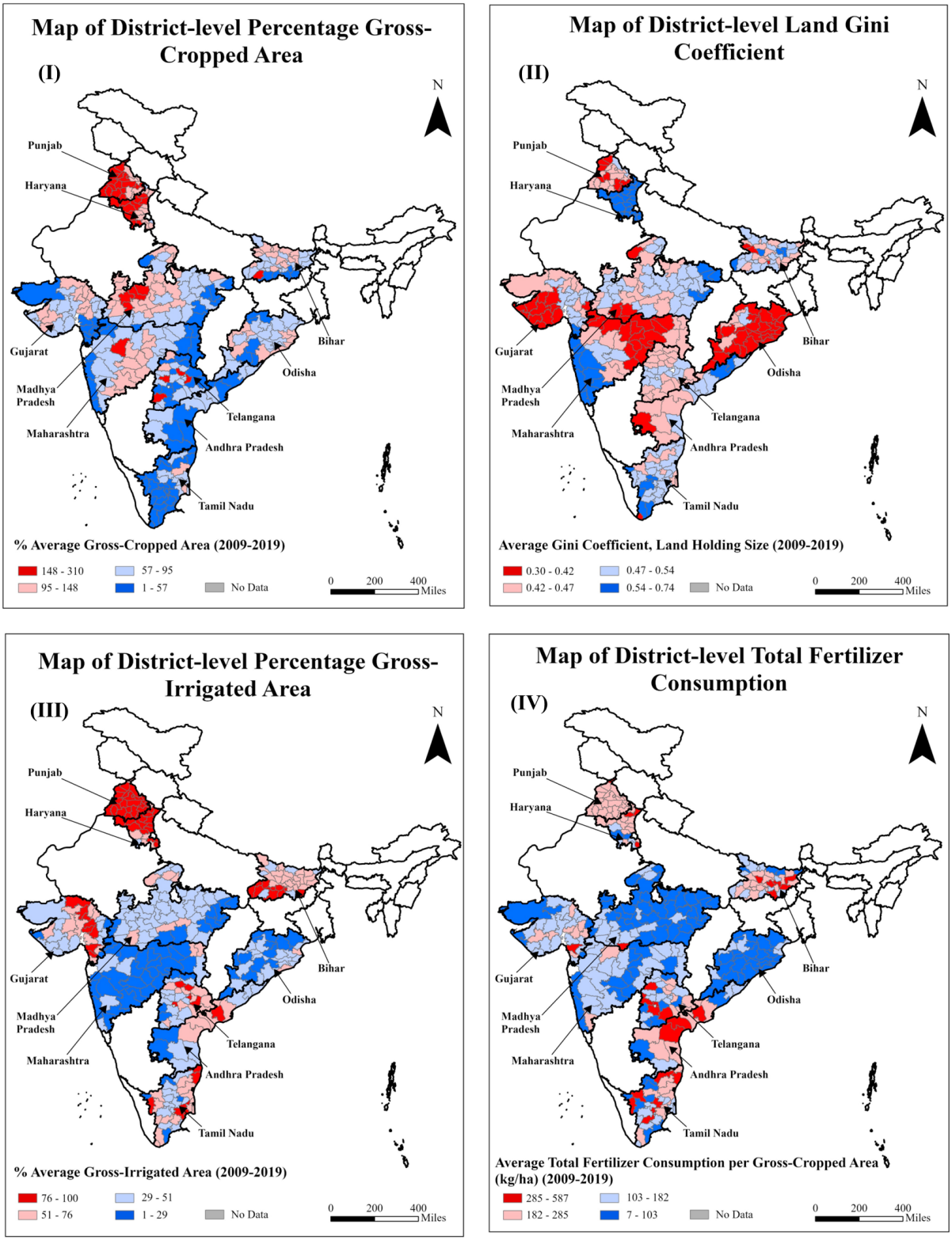
Figure 1: Regional and state-wise distributions of issue types for paddy-related queries.



Note: The average seasonal deviation for each district for the weather variables are 2.42, 2.12, and 3.03, respectively. The average district-level rural population is 78.21% (Table 3).

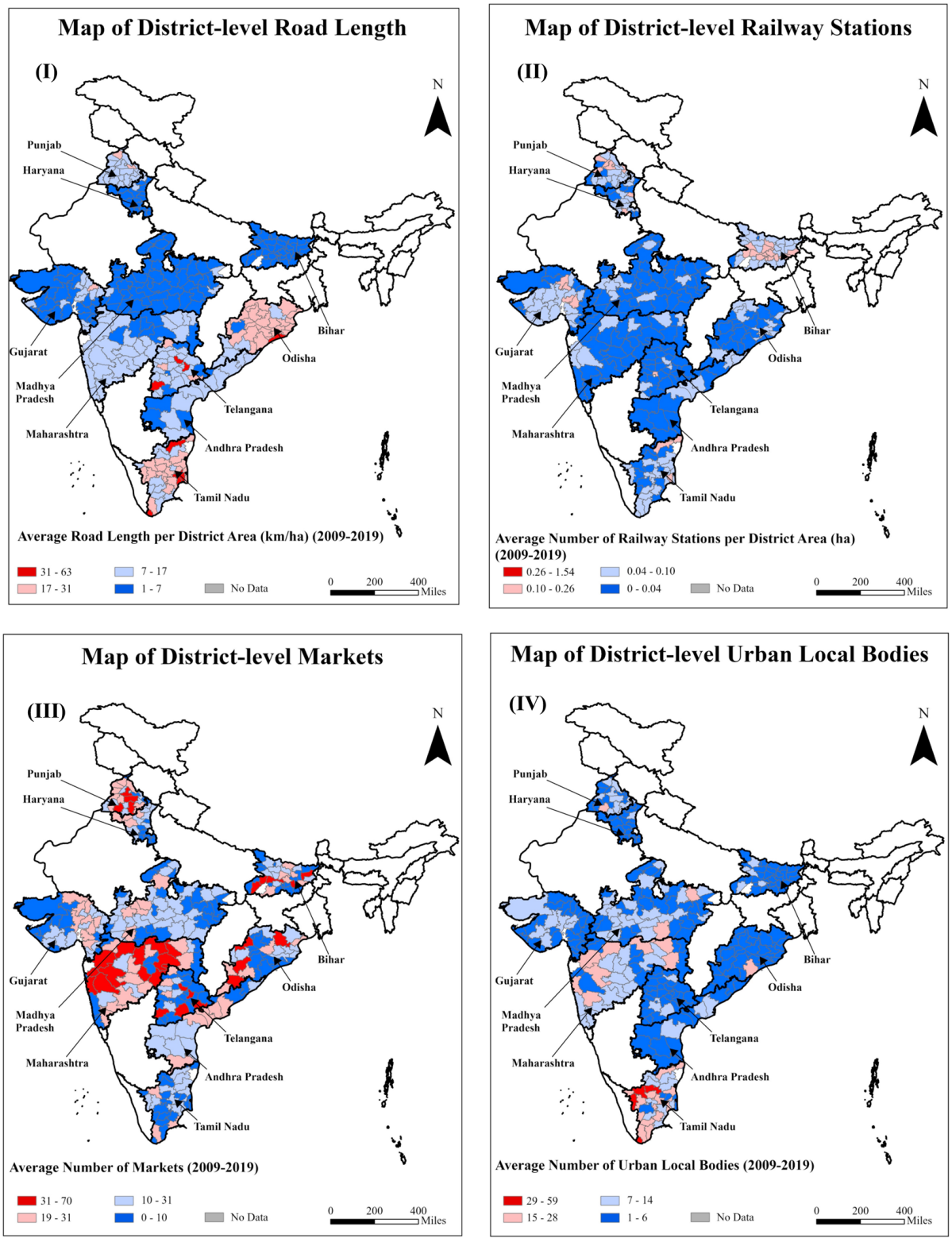
Figure 2: Maps of district-level seasonal maximum temperature, minimum temperature, and rainfall deviations from historical averages and percentage rural population

for selected Indian States (2009-2019)



Note: The average percentage gross-cropped area, Gini coefficient based on landholding size, percentage gross irrigated area and total fertilizer consumption per gross cropped area are 93%, 0.47, 51% and 156.36 kg/hectare, respectively (Table 3).

Figure 3: Maps of district-level distribution of gross cropped area and land Gini coefficient, and gross irrigated area and fertilizer consumption for selected Indian States (2009-2019)

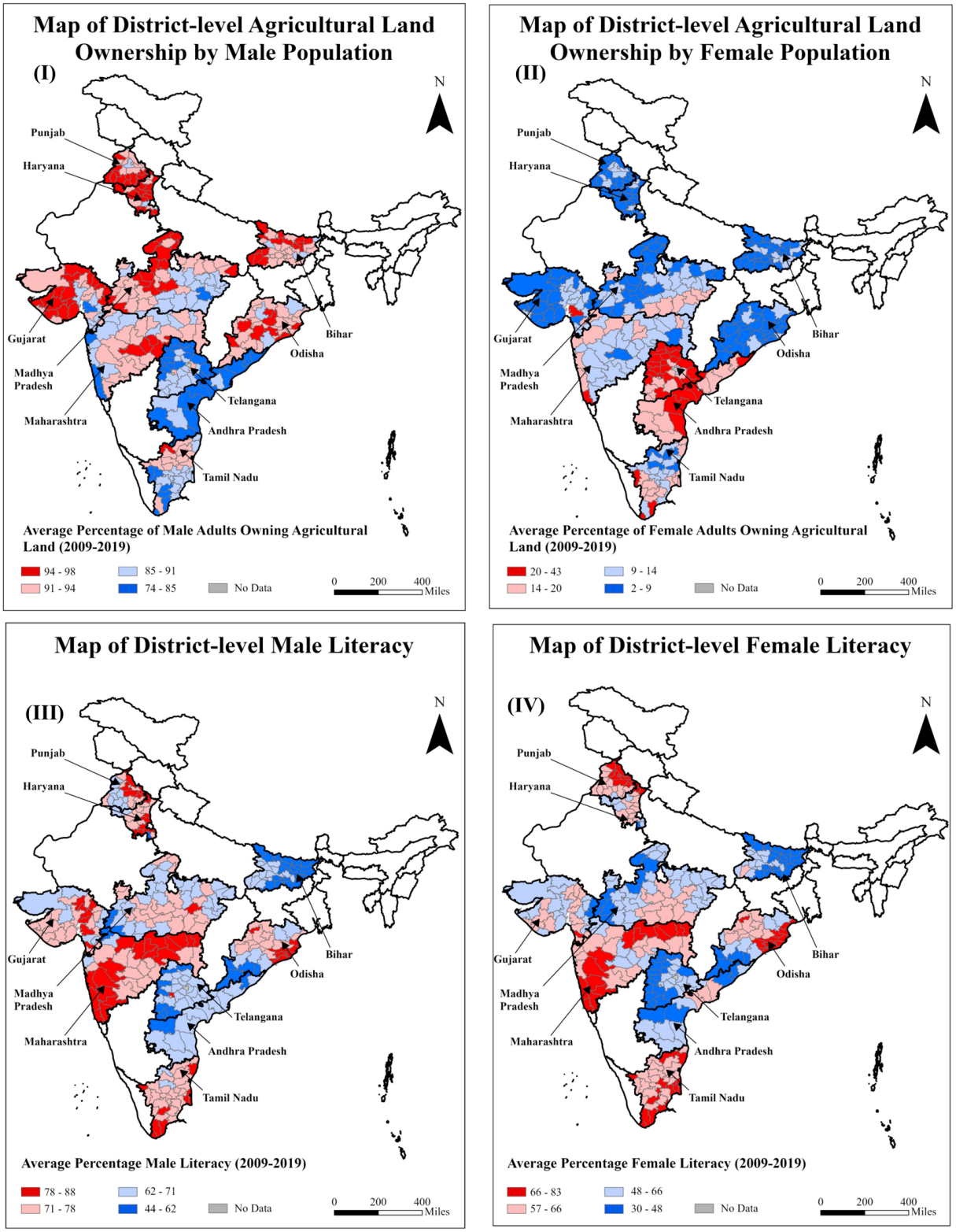


Note: The average road length per district area, number of railway stations per district area, number of markets, and number of urban local bodies are 10.42 km/hectare, 0.05, 15, and 7, respectively (Table 3).

Figure 4: Maps of district-level distribution of road length, railway stations, markets, and urban local bodies for selected Indian States (2009-2019)

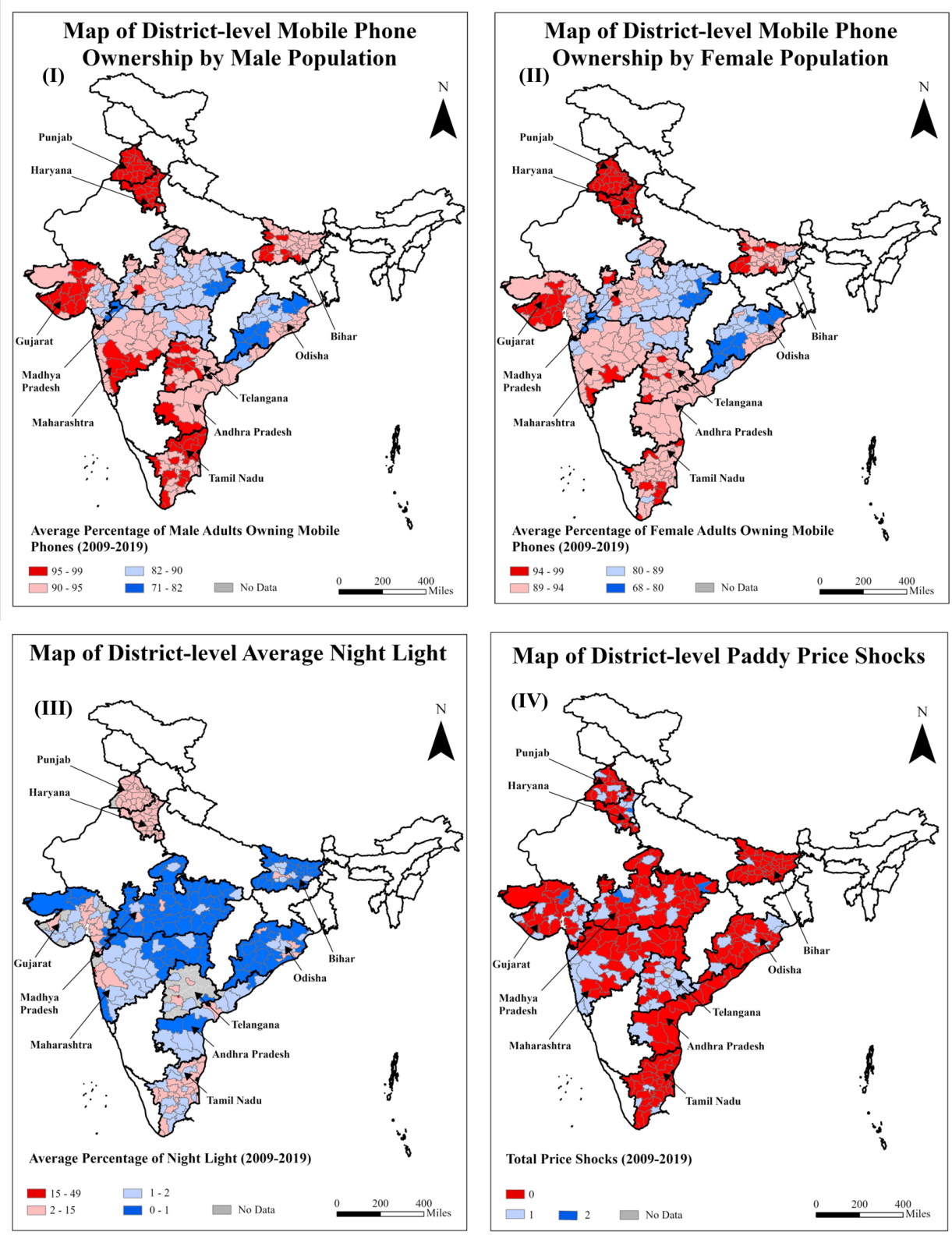


Figure 5: Auto Correlation (ACF) plots of paddy yields and percentage of farmers' queries at KCC for paddy pests and weather by State



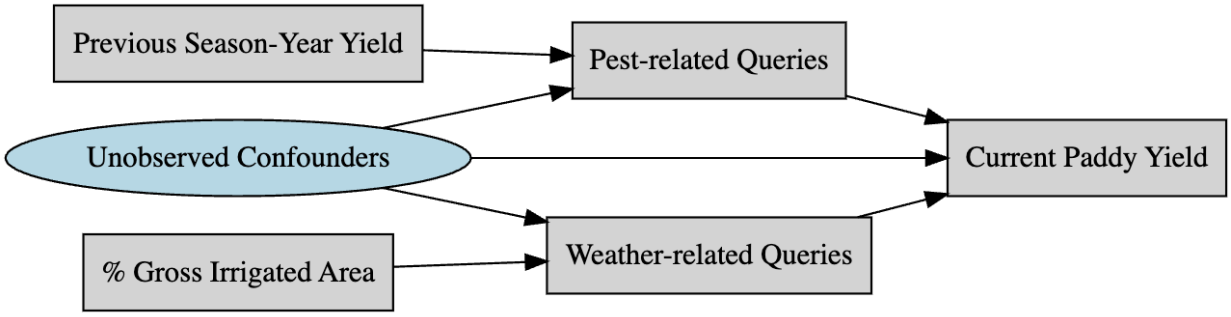
Note: The average agricultural land ownership (male and female) and literacy rate (male and female) are 91.24%, 10.96%, 71.53% and 56.70%, respectively (Table 5).

Figure 6: Maps of district-level distribution of agricultural land ownership and literacy rate by gender for selected Indian States (2009-2019)



Note: The average mobile phone ownership (male and female), night light luminosity, and paddy price shocks are 92.44%, 91.39%, 2.14, and 0.03, respectively (Table 5).

Figure 7: Maps of district-level distribution of gender based mobile phone ownership, night light luminosity, and paddy price shocks (2009-2019)



Note: This diagram illustrates the causal logic of the instrumental variable (IV) approach used in the analysis. The instrumentsprevious seasons paddy yield and percentage of gross irrigated areaare assumed to influence farmers pest- and weather-related query behavior without directly affecting current paddy yields, except through the treatment channels. Unobserved confounders (e.g., farmer distress, local shocks) may affect both query behavior and yield outcomes, motivating the use of IV estimation to isolate exogenous variation in treatment exposure.

Figure 8: Conceptual path diagram of instrumental variable strategy

Table 1: Summary statistics of query types

	Mean	% Share	Median	S.D.	N
Issue type					
Pest	0.33	33%	0.00	0.47	17058437
Weather	0.35	35%	0.00	0.48	17058437
Contact information	0.02	2%	0.00	0.13	17058437
Finance	0.07	7%	0.00	0.25	17058437
Price	0.00	0%	0.00	0.06	17058437
Fertilizer	0.09	9%	0.00	0.28	17058437
Weed	0.03	3%	0.00	0.16	17058437
Input	0.12	12%	0.00	0.33	17058437
Escalated issues	0.00	0%	0.00	0.01	17058437
Other issues	0.18	18%	0.00	0.38	17058437
Crop type					
Paddy	0.08	8%	0.00	0.27	17058437
Wheat	0.04	4%	0.00	0.19	17058437
Maize	0.01	1%	0.00	0.10	17058437
Cotton	0.04	4%	0.00	0.20	17058437
Sugarcane	0.01	1%	0.00	0.09	17058437
Vegetables	0.13	13%	0.00	0.33	17058437
Pulses	0.04	4%	0.00	0.21	17058437
Oilseed	0.01	1%	0.00	0.11	17058437
Livestock	0.01	1%	0.00	0.10	17058437
Fruits	0.04	4%	0.00	0.19	17058437
Millet	0.00	0%	0.00	0.07	17058437
Fodder	0.00	0%	0.00	0.04	17058437

Note: This table provides summary statistics of the dummy variables (0 or 1) categorized under various issues and crop types. The "Other" variable includes issue-related activities beyond weather, pest, contact, finance, price, fertilizer, weed, and input. Additionally, the table presents "Escalated issues," referring to farmers' questions that were to higher officials for further information. The mean of the queries reflects the percentage share of the query issue and crop types. The classification of farmers' queries at KCCs was accomplished using a text analysis algorithm in R (Ott et al., 2011; Naldi, 2019). The algorithm categorized the queries into a dataset comprising individual words associated with respective dummy variables. To ensure accuracy, the words underwent spell-checking, accounted for local languages, and incorporated technical terms related to agriculture (Etoori et al., 2018). Also, as a process of text-mining, we categorized English and non-English high-frequency words. Developed an agronomic data dictionary based on the high-frequency issues and crop types of the KCC conversations. The technical and non-English (i.e., local agronomic languages) are appropriately matched with the issue and crop types. We have randomly checked the classification of the queries under the above-mentioned categories in the agronomic data dictionary, which were 100% accurate. We acknowledge we might have missed a few local agronomic or technical terms. The Text analysis provides an indication of the seriousness of a particular issue affecting agricultural production within districts in India. Analysis of textual data from social media, digital news media, and internet searches provides estimates for important economic indicators (Gentzkow et al., 2019; Antweiler and Frank, 2004; Scott and Varian, 2015).

Table 2: State-wise distribution of weather and pest issue types in farmers' queries with percentages at KCCs

State Name	Weather = 0 (N = 11,078,422)		Weather = 1 (N = 5,980,015)		Total
	Pest = 0 (N = 5,600,612)	Pest = 1 (N = 5,477,810)	Pest = 0 (N = 5,781,129)	Pest = 1 (N = 198,886)	
Punjab	451,082 (23.94%)	518,081 (27.50%)	898,749 (47.70%)	16,190 (0.86%)	1,884,102 (11.04%)
Haryana	454,543 (23.06%)	531,523 (26.97%)	981,101 (49.77%)	3,978 (0.20%)	1,971,145 (11.56%)
Andhra Pradesh	238,766 (31.55%)	265,131 (35.03%)	252,212 (33.32%)	735 (0.10%)	756,844 (4.44%)
Telangana	230,065 (30.68%)	273,151 (36.42%)	245,858 (32.78%)	886 (0.12%)	749,960 (4.40%)
Tamil Nadu	669,491 (48.09%)	521,690 (37.47%)	186,209 (13.38%)	14,795 (1.06%)	1,392,185 (8.16%)
Odisha	312,795 (28.71%)	511,320 (46.93%)	154,230 (14.16%)	111,213 (10.21%)	1,089,558 (6.39%)
Bihar	685,800 (38.79%)	467,028 (26.42%)	613,109 (34.68%)	2,041 (0.12%)	1,767,978 (10.36%)
Gujarat	664,738 (52.31%)	415,889 (32.73%)	169,375 (13.33%)	20,743 (1.63%)	1,270,745 (7.45%)
Maharashtra	1,066,158 (30.29%)	1,329,513 (37.77%)	1,115,513 (31.69%)	8,513 (0.24%)	3,519,697 (20.63%)
Madhya Pradesh	827,174 (31.14%)	858,484 (32.32%)	950,773 (35.79%)	19,792 (0.75%)	2,656,223 (15.57%)
Total	5,600,612 (32.83%)	5,470,810 (32.11%)	5,781,129 (33.89%)	198,886 (1.17%)	17,058,437 (100.00%)

Note: The table presents the state-wise distribution of issue types in farmers' queries recorded at the *Kisan* (farmers) call centers (KCC). The two main issue categories*weather* and *pest*are coded as binary indicators (1 = issue mentioned; 0 = not mentioned). The States represent all major agro-climatic regions of India: North (Punjab, Haryana), South (Andhra Pradesh, Telangana, Tamil Nadu), East (Odisha, Bihar), West (Maharashtra, Gujarat), and Central (Madhya Pradesh).

Table 3: Summary statistics of weather, land, crop yields, and infrastructure variables

	Min	Max	Mean	Median	S.D.	N
Weather variables						
Maximum Temperature$_{(ds)ry}$(°C)	19.44	45.80	34.86	34.50	4.62	19782
Minimum Temperature$_{(ds)ry}$(°C)	3.40	28.00	17.60	19.03	5.29	19782
Rainfall$_{(ds)ry}$ (mm)	0.00	5519.90	360.88	141.40	482.94	19782
TD Max$_{(ds)ry}$ (dummy, 0 or 1)	0.00	1.00	0.04	0.00	0.19	19782
TD Min$_{(ds)ry}$ (dummy, 0 or 1)	0.00	1.00	0.03	0.00	0.18	19782
RD$_{(ds)ry}$ (dummy, 0 or 1)	0.00	1.00	0.06	0.00	0.21	19782
Total district-level weather deviations						
TD Max$_{(ds)ry}$ (dummy, 0 or 1)	0.00	5.00	2.42	3.00	1.23	314
TD Min$_{(ds)ry}$ (dummy, 0 or 1)	0.00	5.00	2.12	2.00	1.19	314
RD$_{(ds)ry}$ (dummy, 0 or 1)	1.00	6.00	3.03	3.00	0.98	314
Land and Input variables						
% Gross Cropped Area (GCA)$_{(ds)y}$	0.06	4.85	0.93	0.83	0.55	3154
Gini coefficient (land holding size)$_{(ds)y}$	0.27	0.75	0.47	0.47	0.07	3154
Avg. landholding size$_{(ds)y}$ (hectare)	0.18	6.86	1.47	1.31	0.94	3154
% Gross Irrigated Area (GIA)$_{(ds)y}$	0.00	1.00	0.51	0.48	0.28	3154
NPK per GCA$_{(ds)y}$ (kg/hectare)	1.63	872.99	156.36	135.82	98.78	3141
Crop yields						
Paddy yield$_{(ds)ry}$ (2009-2019) (t ha^{-1})	0.01	6.64	2.60	2.60	1.23	15956
Paddy yield$_{(ds)ry}$ (2008-2018) (t ha^{-1})	0.01	9.89	2.56	2.55	1.24	15245
Infrastructure variables						
# Markets$_{(ds)y}$	0.00	70.00	15.02	12.00	11.09	3289
# ULB$_{(ds)}$ (i.e., municipalities)	1.00	59.00	6.67	5.00	7.01	3289
Road length per total dist. area$_{(ds)y}$ (km/hectare)	0.53	130.56	10.42	7.88	10.03	3261
# Train stations per total dist. area$_{(ds)}$	0.00	1.54	0.05	0.04	0.10	3261
Demographic variable						
% Rural population$_{(ds)y}$	0.00	96.53	72.56	78.21	19.59	4104

Note: Weather dummy variables: If the observed seasonal temperatures and rainfall deviate by $\pm$ 2 $S.D.$ from the historical mean (2000-19), the indicator variables are coded as 1; otherwise, they are 0.
Two sets of crop yields in analysis; the current year (2009-2019) and the previous year (2008-2018).
All the observations are based on 314 districts across ten selected States of India. Subscripts for district d, State s, season r and year y.

Table 4: State-wise monthly distribution of seasons

Sl no.	Region	State Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	North	Punjab	Winter			Summer			Monsoon			Post-monsoon		Winter
2	North	Haryana	Winter		Summer				Monsoon			Post-monsoon		Winter
3	West	Gujarat	Winter		Summer				Monsoon			Post-monsoon		Winter
4	West	Maharashtra	Winter		Summer			Monsoon				Post-monsoon		Winter
5	Central	Madhya Pradesh	Winter		Summer				Monsoon			Post-monsoon		Winter
6	East	Bihar	Winter			Summer			Monsoon			Post-monsoon		Winter
7	East	Odisha	Winter		Summer				Monsoon			Post-monsoon		Winter
8	South	Telangana	Winter		Summer				Monsoon			Post-monsoon		Winter
9	South	Andhra Pradesh	Winter		Summer			Monsoon				Post-monsoon		Winter
10	South	Tamil Nadu	Winter		Summer				Monsoon					Winter

Note: The mapping of seasons with respect to the month and the administrative regions is based on information from the government and other reliable websites of each State.

Table 5: Summary statistics of robustness and placebo checks variables

	Min	Max	Mean	Median	S.D.	N
Weather z-score Variables						
Maximum Temperature$_{(ds)ry}$	-3.21	2.30	0.00	-0.08	1.00	19782
Minimum Temperature$_{(ds)ry}$	-2.65	1.93	0.00	0.27	1.00	19782
Rainfall$_{(ds)ry}$	-0.74	9.66	0.00	-0.47	1.00	19782
DHS variables						
% Agricultural Landholding (Male)$_{(ds)y}$	67.13	99.12	91.24	92.23	4.50	575
% Agricultural Landholding (Female)$_{(ds)y}$	0.88	67.13	10.96	9.51	6.02	575
% Male Literate$_{(ds)y}$	40.19	89.09	71.53	72.23	8.02	575
% Female Literate$_{(ds)y}$	25.70	85.57	56.70	56.69	9.77	575
% Mobile Phone Ownership (Male)$_{(ds)y}$	63.16	100.00	92.44	93.98	5.82	575
% Mobile Phone Ownership (Female)$_{(ds)y}$	60.65	99.56	91.39	92.78	6.22	575
Night Light and Paddy Price Shocks						
Night Light Luminosity$_{(ds)y}$ (nW/cm^2/sr)	0.02	63.00	2.14	0.70	5.41	5225
Paddy Price Shocks$_{(ds)y}$(dummy, 0 or 1)	0.00	1.00	0.03	0.00	0.16	3245
Crop yields						
Pulses yield$_{(ds)ry}$ (2009-2019) (t ha^{-1})	0.01	6.94	0.70	8.4	0.44	41742
Wheat yield$_{(ds)ry}$ (2009-2019) (t ha^{-1})	0.18	5.8	2.96	2.86	1.25	6232

Note: Price shock dummy variable: If the observed yearly price deviate by $\pm$ 2 $S.D.$, from the mean (2009-19) the indicator variable is coded as 1; otherwise, they are 0. Pulses group include the following crops: Pigeon Pea, Cowpea, Green Gram, Black Gram, Lenti, Grass Pea, Moth Bean, Peas and Beans, Other monsoon pulses, Other winter pluses, Other summer pulses. All the observations are based on 314 districts across ten selected States of India. Subscripts for district d, State s, season r and year y. nW/cm^2/sr: nanowatts per square centimeter per steradian.

Table 6: Paddy yields and farmers' queries

Treatment Variables	Dependent Variable: Paddy yields (t ha^{-1})					
	OLS			IV		
	Pest Only	Weather Only	Both	Pest Only	Weather Only	Both
% Pest Queries$_{(ds)ry}$	0.001		0.002**	−0.329		−0.029**
	(0.001)		(0.001)	(1.190)		(0.014)
% Weather Queries$_{(ds)ry}$		0.007***	0.007***		0.021***	0.008*
		(0.001)	(0.001)		(0.003)	(0.005)
District × Season f.e.	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓
Observations	8,828	8,828	8,828	8,828	8,828	8,828
Weak Instrument statistic (pest)				97.150***		7.982***
p-value				($< 2 \times 10^{-16}$)		(0.0004)
Weak Instrument statistic (weather)					89.300***	103.220***
p-value					($< 2 \times 10^{-16}$)	(2×10^{-16})
Wu–Hausman statistic				7.336***	16.240***	9.426***
p-value				(4.35×10^{-5})	(5.75×10^{-5})	(8.37×10^{-5})
Wald statistic				38.070***	23.880***	13.740***
p-value				($< 2 \times 10^{-16}$)	($< 2 \times 10^{-16}$)	($< 2 \times 10^{-16}$)

Note: The analysis involves district-level data from 10 states from 2009 to 2019. The KCC data in the analysis are related to the paddy crop for pest and weather queries. The table reports OLS estimates (Columns 1 to 3) and 2SLS/IV (Columns 4 to 6) on the effect of percentage seasonal queries on pest and weather (as treatment variables) at the district level on seasonal paddy yield (as outcome variable) based on equation 1. All 2SLS/IV specifications include the instrumental variables; the previous year's paddy yield for pest queries and the % gross irrigated area for weather queries. All regression equations include district × season-fixed effects, baseline controls, and additional controls. The baseline controls include seasonal recorded maximum and minimum temperatures and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Columns under heading "both" represent both pest and weather related queries are used as treatments. Standard errors are clustered at the district × season level (in parentheses). Wu–Hausman statistics test endogeneity; weak-instrument statistics evaluate instrument strength; Wald statistics test over-identifying restrictions. Subscripts: district d, state s, season r, year y. Significance: *p< 0.1, **p< 0.05, ***p< 0.01.

Table 7: Robustness checks: Paddy yields and farmers' queries (Driscoll-Kraay clustered standard errors and price shock controls)

Treatment Variables	Dependent Variable: Paddy yields (t ha^{-1})					
	(OLS)			(IV)		
	Pest only	Weather only	Both	(Pest only)	Weather only	Both
	Driscoll-Kraay clustered standard error estimates					
% Pest Queries$_{(ds)ry}$	0.001		0.002*	−0.329		−0.029**
	(0.001)		(0.001)	(1.447)		(0.017)
% Weather Queries$_{(ds)ry}$		0.007***	0.007***		0.021***	0.008
		(0.001)	(0.001)		(0.005)	(0.006)
District × Season f.e.	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓
	Estimates including price shocks					
% Pest Queries$_{(ds)ry}$	0.001		0.002*	−0.351		−0.029**
	(0.001)		(0.001)	(1.378)		(0.015)
% Weather Queries$_{(ds)ry}$		0.007***	0.007***		0.021***	0.007
		(0.001)	(0.001)		(0.003)	(0.005)
District × Season f.e.	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓
Price shocks	✓	✓	✓	✓	✓	✓
Observations	8,828	8,828	8,828	8,828	8,828	8,828

Note: The analysis involves district-level data from 10 states from 2009 to 2019. The KCC data in the analysis are related to the paddy crop for pest and weather queries. The table reports OLS estimates (Columns 1 to 3) and 2SLS/IV (Columns 4 to 6) on the effect of percentage seasonal queries on pest and weather (as treatment variables) at the district level on seasonal paddy yield (as outcome variable) based on equation 1. All 2SLS/IV specifications include the instrumental variables; the previous year's paddy yield for pest queries and the % gross irrigated area for weather queries. All regression equations include district × season-fixed effects, baseline controls, and additional controls. The baseline controls include seasonal recorded maximum and minimum temperatures and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Columns under heading "both" represent both pest and weather related queries are used as treatments. Standard errors are clustered at the district × season level (in parentheses). Wu–Hausman statistics test endogeneity; weak-instrument statistics evaluate instrument strength; Wald statistics test over-identifying restrictions. Subscripts: district d, state s, season r, year y. Significance: *p< 0.1, **p< 0.05, ***p< 0.01.

Table 8: Robustness checks: Paddy yields and farmers' queries (Extended fixed effects and controls)

Treatment Variables	Dependent Variable: Paddy yields (t ha^{-1})										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
% Pest	-0.029^{*}	-0.033^{*}	-0.032^{*}	-0.029^{**}	-0.031^{**}	-0.031^{**}	-0.031^{**}	-0.031^{**}	-0.030^{**}	-0.030^{*}	-0.032^{**}
Queries$_{(ds)ry}$	(0.016)	(0.017)	(0.017)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)	(0.015)
% Weather	0.008	0.007	0.013	0.007*	0.006	0.007	0.007	0.007	0.008	0.008	0.010
Queries$_{(ds)ry}$	(0.006)	(0.006)	(0.010)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Observations	8,828	8,828	8,828	8,828	8,828	8,828	8,828	8,828	8,828	8,828	8,828
Wu-Hausmann statistic	9.426***	9.921***	5.720**	9.123***	8.415***	8.315***	6.189***	8.980***	9.188***	8.295***	8.707***
p-value	(0.0001)	(0.0001)	(0.0033)	(0.0001)	(0.0002)	(0.0003)	(0.002)	(0.0001)	(0.0001)	(0.0003)	(0.0002)
Fixed Effects											
District ×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Agroclimatic	✓	×	×	×	×	×	×	×	×	×	×
Landholding type	×	✓	×	×	×	×	×	×	×	×	×
Mobile Boom	×	×	✓	×	×	×	×	×	×	×	×
Controls											
Baseline	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
Additional	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Weather z-scores	×	×	×	✓	×	×	×	×	×	×	×
Mobile (male)	×	×	×	×	✓	×	×	×	×	×	×
Mobile (female)	×	×	×	×	×	✓	×	×	×	×	×
Night Light	×	×	×	×	×	×	✓	×	×	×	×
Land (male)	×	×	×	×	×	×	×	✓	×	×	×
Land (female)	×	×	×	×	×	×	×	×	✓	×	×
Literacy (male)	×	×	×	×	×	×	×	×	×	✓	×
Literacy (female)	×	×	×	×	×	×	×	×	×	×	✓

Note: The table reports 2SLS-IV estimates on the effect of percentage seasonal queries on pest and weather at the district level on paddy yields. All specifications include the instrumental variables; the previous year's paddy yield for pest queries and the % gross irrigated area for weather queries. All regression equations include district × season-fixed effects, baseline controls, and additional controls. The baseline controls include; seasonal recorded maximum and minimum temperatures, and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include; the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Standard errors are clustered at the district × season level reported in parenthesis. Each column represent additional fixed-effects or controls. Columns (1) to (3) - Agro-climatic zone, landholding size type (i.e., Small (0-2 ha), Medium (2-10 ha), and Large (>10 ha)) and mobile boom (i.e., indicator variable: post-2012 coded as 1 other wise 0, following World Bank data) fiexed effects, respectively. The columns (4) to (11) represent control variables includes weather z-scores instead of weather deviation variables under baseline controls, mobile phone ownerships (male and female), night light luminosity, agricultural land ownerships (male and female), literacy rates (male and female) . Subscripts for district d, State s, season r and year y. Significance level: *p<0.1; **p<0.05; ***p<0.01.

Table 9: Robustness checks: Paddy yields and farmers' queries (Alternate IVs, additional IVs and seasons)

Treatment Variables	Dependent Variable: Paddy yields (t ha^{-1})						
	RGY	Log-Yields	Alternate IVs	Additional IVs	Monsoon	Winter	Summer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Pest Queries$_{(ds)ry}$	-0.005^{*}	-0.005	-0.006	-0.010^{*}	-0.048	-0.019^{***}	-0.016
	(0.003)	(0.008)	(0.010)	(0.006)	(0.095)	(0.007)	(0.045)
% Weather Queries$_{(ds)ry}$	0.002^{**}	0.004	0.012^{***}	0.012^{***}	0.001	0.012	-0.004
	(0.001)	(0.003)	(0.003)	(0.003)	(0.034)	(0.012)	(0.017)
Observations	8,828	8,828	8,828	8,828	8,828	8,828	8,828
Wu-Hausmann statistic	10.564^{***}	0.961	2.737^{*}	8.080^{***}	12.88^{***}	3.744^{*}	4.184^{*}
p-value	(0.00003)	(0.3828)	(0.0650)	(0.0003)	(0.0001)	(0.0251)	(0.0157)
Hansen J statistic				0.015			
p-value				(0.9019)			
District $\times$ Season f.e.	✓	✓	✓	✓	$\times$	$\times$	$\times$
District f.e.	$\times$	$\times$	$\times$	$\times$	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓

Note: The table reports 2SLS-IV estimates on the effect of percentage seasonal queries on pest and weather at the district level on paddy yields. All specifications include the instrumental variables; the previous year's paddy yield for pest queries and the % gross irrigated area for weather queries. All regression equations include district $\times$ season-fixed effects (except season-level analysis in columns (4) to (6)), baseline controls, and additional controls. The baseline controls include; seasonal recorded maximum and minimum temperatures, and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include; the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Standard errors are clustered at the district $\times$ season level reported in parenthesis. Each columns have different specifications.Column (1): Outcome variable as Relative grain yield (RGY) of paddy Column (2): Outcome variable as log paddy yields, column (3): Alternate IV definitions (average of 1-year and 2-year lag yields instruments as IV for pest-related queries and 5-year rolling CAGR % gross irrigated area as IV for weather-related queries), column (4): Additional IV for weather-related queries % gross irrigated area and 5-year rolling CAGR % gross irrigated area, columns (5)-(7) present estimated treatment estimates based on seasons (monsoon, wither and summer) with district fixed-effect. Subscripts for district d, State s, season r and year y. Significance level: *p<0.1; **p<0.05; ***p<0.01.

Online Appendix

A *Kisan* (Farmers) Call Center

The *Kisan* (farmers) Call Center initiative was launched by the Indian government's Ministry of Agriculture and Farmers Welfare on January 21, 2004, with the primary objective of providing valuable assistance to farmers in their agricultural endeavors. This scheme aims to harness the potential of digital technologies in agriculture to empower farmers with crucial information and guidance. The farmers' call centers operate in 22 vernacular languages and are located across 21 different regions in India. They offer telephonic support and address farmers' queries from 6.00 AM to 10.00 PM, seven days a week. The farm tele-advisors are well-equipped, as they are agriculture graduates with training in various allied disciplines, such as Horticulture, Animal Husbandry, Fisheries, Poultry, Bee-Keeping, Sericulture, Aquaculture, Agricultural Engineering, Agricultural Marketing, and Bio-Technology. Their expertise ensures that farmers receive reliable and relevant advice to enhance their agricultural practices.

The KCC facilities are easily accessible through a common toll-free number, 1800-180-1551, from both mobile phones and landline telephones in India. When farmers seek information on their farming activities, they provide their name, phone number, and village name to the tele-advisors at the KCC system. These knowledgeable tele-advisors address the farmers' queries, and if necessary, they can escalate the queries to higher-level experts. Only a small percentage of queries require escalation, as indicated in Table 1[31]. The level-II officers, senior tele-advisors, handle the queries on the same call. If the query remains unresolved, it is forwarded to level-III officers, experts in their respective fields. These experts are specialists from state and central government institutions engaged in research and extension activities and are part of the KCC network nationwide. Farmers receive calls from subject matter specialists for queries within 72 hours. The KCC operations are regularly monitored and reviewed by nodal institutions at various levels to ensure the call center's proper functioning. Moreover, the KCC data on farmers' queries are available on the ministry's data web portal at both month and district levels, facilitating access to valuable agricultural information.

The KCC data has been accessed through the *Kisan* Call Center web portal for the following

[31] There were 1573 queries out of 17 million queries, the lowest among the identified query types.

States of India; North (Punjab, Haryana), South (Andhra Pradesh, Telangana, Tamil Nadu), East (Odisha, Bihar), West (Maharashtra, Gujarat), Central (Madhya Pradesh). The KCC database can be accessed through the open data platform of the Ministry of Agriculture by web crawling and integrated with other databases through a process-based methodology (Figure A.1).

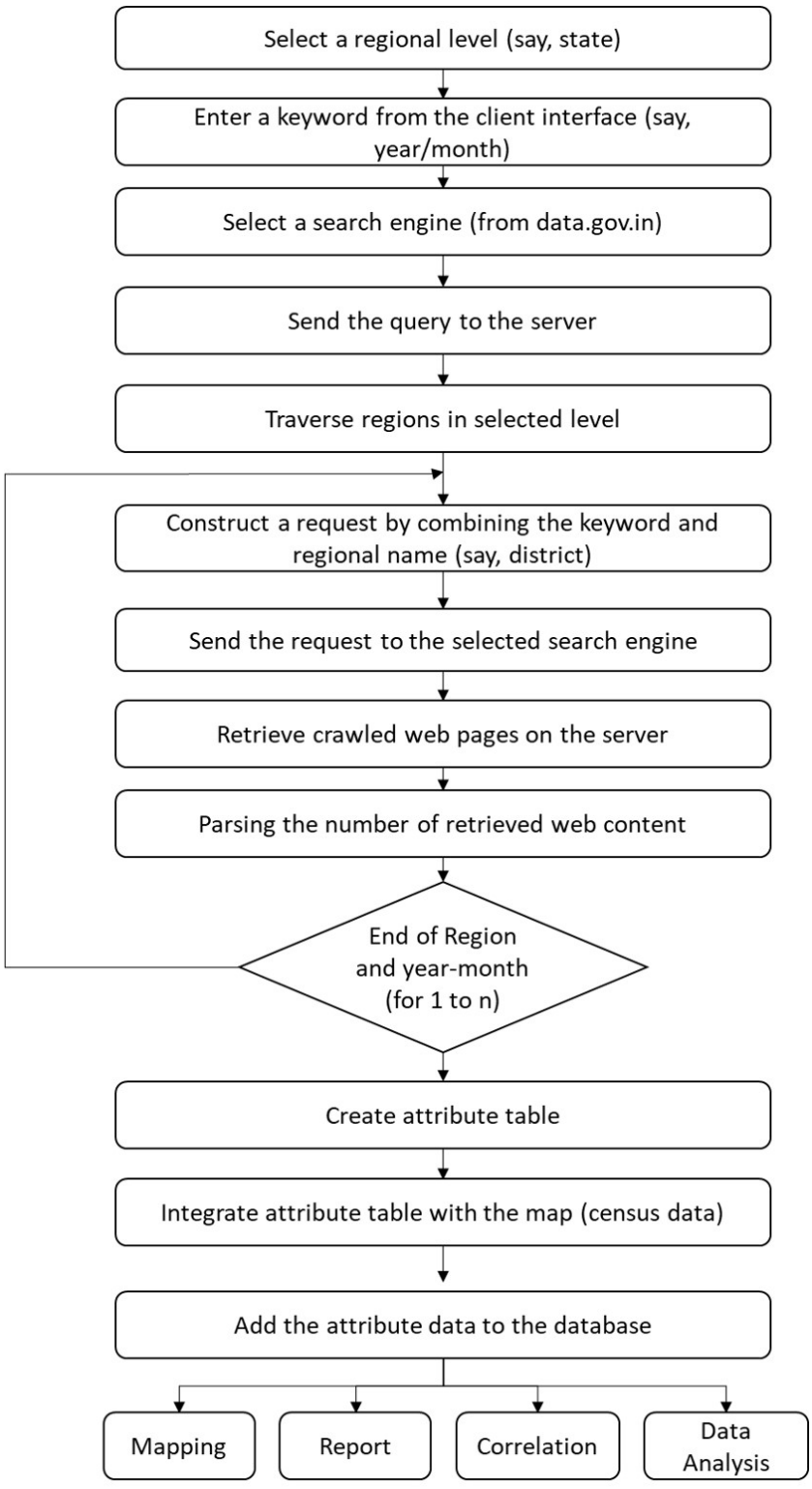
A sample of KCC queries is tabulated under Table A.1. The queries and answers are reported concerning the call's time and the farmer's location (block, district, and state). Text analysis has been performed on the KCC queries to classify the queries. The queries are grouped by the problems that farmers require information (weather, pest, weed, finance, price, fertilizer, other inputs, contact information, and queries)[32] for different crops like paddy, wheat, maize, cotton, sugarcane, vegetables, pulses, oilseeds, livestock, fruit, millet, and fodder.

The volume of queries combining all the years (2009-2019) for ten States is over 17 million, Table A.2. The volume of queries reported at KCCs has increased over the years, signifying the importance of information in agriculture. The scientific information provided by the KCCs is expected to significantly influence the crop yield, as the growth of incidence of queries is significant from 2009 to 2019 (CAGR of total queries for ten States is 31%).

The farmers' queries provide valuable insights into regional and seasonal agricultural issues. Maharashtra contributed 20.63%, and Madhya Pradesh contributed 15.57% of the reported queries. Analyzing the farmers' queries at the seasonal level highlights a pattern for the incidence of queries Table A.3. The call volumes are highly correlated with the growing season of key crops within selected Indian States[33].

[32]Finance: Queries are related to direct benefit transfer, *Kisan* (farmer) credit card, loan, subsidies, market and schemes; Price: Queries related to crop prices, minimum support price; Inputs: Queries related to seed varieties, cropping practices, and irrigation; Contact: Queries related to contact information to Krishi Vigyan Kendra (agricultural extension center), State government

[33]The farmers' call centers in Punjab and Haryana saw the highest call volumes between June and September. Similarly, in Odisha, the majority of KCC queries occurred from July to October, while in Bihar, queries were more prevalent from June to September. The states of Andhra Pradesh, Telangana, and Tamil Nadu reported the highest number of KCC queries from August to December. In the western and central regions of India, encompassing Gujarat, Maharashtra, and Madhya Pradesh, the peak KCC query period was from June to September. This seasonal pattern of queries reflects the cropping seasons in these regions.



Source: Modified flowchart from She et al. (2017).

Figure A.1: Procedure for the integration of Internet data and census data

Table A.1: Sample of Farmers' Queries at *Kisan* Call Center (KCC)

Season	Crop	Query Text	KCC Answer	State	District	Block	Created On
Kharif	Others	Problem related to DBT	For farmer registration: www.krishimis.in	Bihar	Darbhanga	Ghyanshyampur	8/3/2018 09:13
Kharif	Wheat	Improved varieties of wheat crop?	WH-711, PBW-343, WH-147, UP-2338, RAJ-3765	Haryana	Bhiwani	Tosham	12/1/2012 06:09
Jayad	Brinjal	Fruit & Shoot Borer	Spray of Indoxacarb	Odisha	Anugul	Chhendipada	7/23/2013 16:14
Kharif	Paddy (Dhan)	SRI method	Detail POP explained	Odisha	Ganjam	Bhanjanagar	8/3/2018 18:14
Jayad	Others	Weather of LDH?	100% chance of rain at night	Punjab	Ludhiana	Doraha	7/13/2012 08:03

Source: Kisan Call Center (KCC) database.

Table A.2: Year- and State-wise Volume of Queries at KCCs with Percentages

State Name	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Punjab	40,819 (2.17%)	37,742 (2.00%)	37,744 (2.00%)	83,754 (4.45%)	132,651 (7.04%)	224,578 (11.92%)	249,859 (13.26%)	239,514 (12.71%)	268,232 (14.24%)	214,360 (11.38%)	354,849 (18.83%)	1,884,102 (11.04%)
Haryana	30,686 (1.56%)	30,868 (1.57%)	27,235 (1.38%)	68,009 (3.45%)	148,701 (7.54%)	203,840 (10.34%)	259,190 (13.15%)	282,502 (14.33%)	310,459 (15.75%)	249,135 (12.64%)	360,520 (18.29%)	1,971,145 (11.56%)
Odisha	12,811 (1.18%)	18,872 (1.73%)	13,018 (1.19%)	14,342 (1.32%)	75,663 (6.94%)	147,823 (13.57%)	202,926 (18.62%)	182,456 (16.75%)	162,160 (14.88%)	119,864 (11.00%)	139,623 (12.81%)	1,089,558 (6.39%)
Bihar	14,417 (0.82%)	15,108 (0.85%)	15,153 (0.86%)	26,499 (1.50%)	106,808 (6.04%)	157,363 (8.90%)	217,520 (12.30%)	311,831 (17.64%)	257,773 (14.58%)	203,261 (11.50%)	442,245 (25.01%)	1,767,978 (10.36%)
Andhra Pradesh	5,193 (0.69%)	6,000 (0.79%)	10,925 (1.44%)	24,567 (3.25%)	60,743 (8.03%)	72,720 (9.61%)	127,152 (16.80%)	135,704 (17.93%)	101,674 (13.43%)	78,209 (10.33%)	133,957 (17.70%)	756,844 (4.44%)
Telangana	2,979 (0.40%)	5,583 (0.74%)	20,729 (2.76%)	40,251 (5.37%)	60,490 (8.07%)	70,257 (9.37%)	117,358 (15.65%)	148,041 (19.74%)	138,579 (18.48%)	68,663 (9.16%)	77,030 (10.27%)	749,960 (4.40%)
Tamil Nadu	13,398 (0.96%)	31,792 (2.28%)	25,845 (1.86%)	50,616 (3.64%)	143,135 (10.28%)	167,851 (12.06%)	193,641 (13.91%)	235,625 (16.92%)	214,354 (15.40%)	164,256 (11.80%)	151,672 (10.89%)	1,392,185 (8.16%)
Gujarat	41,385 (3.26%)	48,631 (3.83%)	51,646 (4.06%)	80,791 (6.36%)	144,282 (11.35%)	161,410 (12.70%)	169,265 (13.32%)	140,468 (11.05%)	128,980 (10.15%)	108,608 (8.55%)	195,279 (15.37%)	1,270,745 (7.45%)
Maharashtra	49,241 (1.40%)	41,078 (1.17%)	25,822 (0.73%)	81,421 (2.31%)	362,426 (10.30%)	444,272 (12.62%)	597,516 (16.98%)	587,981 (16.71%)	517,081 (14.69%)	365,306 (10.38%)	447,553 (12.72%)	3,519,697 (20.63%)
Madhya Pradesh	62,872 (2.37%)	32,786 (1.23%)	31,610 (1.19%)	102,527 (3.86%)	126,276 (4.75%)	309,112 (11.64%)	395,462 (14.89%)	451,413 (16.99%)	451,799 (17.01%)	299,649 (11.28%)	392,717 (14.78%)	2,656,223 (15.57%)
Total	273,801 (1.61%)	268,460 (1.57%)	259,727 (1.52%)	572,777 (3.36%)	1,361,175 (7.98%)	1,959,226 (11.49%)	2,529,889 (14.83%)	2,715,535 (15.92%)	2,551,091 (14.96%)	1,871,311 (10.97%)	2,695,445 (15.80%)	17,058,437 (100.00%)

Table A.3: Month- and State-wise Volume of Queries with Percentages at KCCs (2009–2019)

State Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Punjab	121527 (6.45%)	111592 (5.92%)	132697 (7.04%)	124138 (6.59%)	166776 (8.85%)	181699 (9.64%)	227180 (12.06%)	199082 (10.57%)	200221 (10.63%)	141980 (7.54%)	141655 (7.52%)	135555 (7.19%)	1884102 (11.04%)
Haryana	135043 (6.85%)	117390 (5.96%)	126327 (6.41%)	140073 (7.11%)	175156 (8.89%)	193805 (9.83%)	235035 (11.92%)	221170 (11.22%)	191791 (9.73%)	151555 (7.69%)	141876 (7.20%)	141924 (7.20%)	1971145 (11.56%)
Odisha	73172 (6.72%)	78502 (7.20%)	98247 (9.02%)	68329 (6.27%)	60493 (5.55%)	82262 (7.55%)	109155 (10.02%)	114677 (10.53%)	112775 (10.35%)	115710 (10.62%)	91966 (8.44%)	84270 (7.73%)	1089558 (6.39%)
Bihar	127050 (7.19%)	127224 (7.20%)	147272 (8.33%)	146777 (8.30%)	133586 (7.56%)	161424 (9.13%)	175890 (9.95%)	168649 (9.54%)	162690 (9.20%)	139185 (7.87%)	133608 (7.56%)	144623 (8.18%)	1767978 (10.36%)
Andhra Pradesh	67359 (8.90%)	53057 (7.01%)	43443 (5.74%)	39799 (5.26%)	40964 (5.41%)	50116 (6.62%)	62313 (8.23%)	65954 (8.71%)	73903 (9.76%)	85240 (11.26%)	93085 (12.30%)	81611 (10.78%)	756844 (4.44%)
Telangana	55744 (7.43%)	45669 (6.09%)	38771 (5.17%)	35060 (4.67%)	33411 (4.46%)	60241 (8.03%)	74887 (9.99%)	84409 (11.26%)	91464 (12.20%)	85283 (11.37%)	77210 (10.30%)	67811 (9.04%)	749960 (4.40%)
Tamil Nadu	118692 (8.53%)	104150 (7.48%)	108798 (7.81%)	100721 (7.23%)	102527 (7.36%)	99827 (7.17%)	111184 (7.99%)	107683 (7.73%)	112018 (8.05%)	129351 (9.29%)	148054 (10.63%)	149180 (10.72%)	1392185 (8.16%)
Gujarat	107240 (8.44%)	88572 (6.97%)	80617 (6.34%)	70171 (5.52%)	68886 (5.42%)	102912 (8.10%)	152114 (11.97%)	146148 (11.50%)	137187 (10.80%)	105949 (8.34%)	100665 (7.92%)	110284 (8.68%)	1270745 (7.45%)
Maharashtra	245238 (6.97%)	226597 (6.44%)	273138 (7.76%)	247056 (7.02%)	241363 (6.86%)	330500 (9.39%)	385230 (10.94%)	363254 (10.32%)	345449 (9.81%)	291043 (8.27%)	281999 (8.01%)	288830 (8.21%)	3519697 (20.63%)
Madhya Pradesh	176290 (6.64%)	160106 (6.03%)	178407 (6.72%)	173554 (6.53%)	174120 (6.56%)	262939 (9.90%)	316796 (11.93%)	279920 (10.54%)	262606 (9.89%)	232129 (8.74%)	210643 (7.93%)	228713 (8.61%)	2656223 (15.57%)
Total	1227355 (7.20%)	1112859 (6.52%)	1227717 (7.20%)	1145678 (6.72%)	1197282 (7.02%)	1525725 (8.94%)	1849784 (10.84%)	1750946 (10.26%)	1690104 (9.91%)	1477425 (8.66%)	1420761 (8.33%)	1432801 (8.40%)	17058437 (100.00%)

Note: Cells with monthly shares above 9% are highlighted in green.

B Additional two-stage least squares (2SLS/IV) regression equations and First-stage regression results

In addition to the main regression equation (2.3.4), we regress the current crop yields with two treatment variables ($Queries_pest_{(ds)ry}$ and $Queries_weather_{(ds)ry}$) individually (i.e., equations (B.1) and (B.2)) to assess the coefficient estimates' consistency.

$$\begin{aligned} Yield_{(ds)ry} =& \alpha + \beta_1 Queries_pest_{(ds)ry} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \eta_{(ds)} \times \psi_r + \epsilon_{(ds)ry} \end{aligned} \tag{B.1}$$

$$\begin{aligned} Yield_{(ds)ry} =& \alpha + \beta_2 Queries_weather_{(ds)ry} + \gamma_1 X_{(ds)ry} + \gamma_2 Y_{(ds)ry} \\ &+ \eta_{(ds)} + \psi_r + \eta_{(ds)} \times \psi_r + \epsilon_{(ds)ry} \end{aligned} \tag{B.2}$$

The coefficients β_1 and β_2 capture the effect of changes in the percentage of seasonal queries on pests and weather at the district level for paddy crop on the current year's yields ($Yield_{(ds)ry}$). The estimations are instrumented with previous years' yields ($Yield_{(ds)r(y-1)}$) and the percentage of gross irrigated area ($Irri_{(ds)ry}$) for percentage seasonal queries on pests and weather at the district level, respectively. The variables $X_{(ds)ry}$ and $Y_{(ds)ry}$ represent baseline controls for the agricultural environment and additional controls for infrastructure support, respectively. District (η_d) $\times$ season (ψ_r) fixed effects are added for each analysis, and the standard errors are clustered at the district $\times$ season level. The results are tabulated in columns (4) and (5) of Table 6 for paddy crop.

The first-stage regression results, following equations (1) and (2), are presented in Table B.5 for paddy. Columns (1) and (3) contain estimates with baseline controls only, while columns (2) and (4) include estimates with both baseline and additional controls. Notably, the instrumental variables demonstrate significant correlation with their respective instruments: the percentage of seasonal queries on pests at the district level for previous years' yields and the percentage of seasonal queries on weather at the district level for the percentage of gross irrigated area. For the paddy crop, both instruments exhibit positive correlations.

Table B.1: Comparison of empirical strategies in ICT-based agricultural information studies

Study	Empirical design	Context / Intervention	Main identification strategy and findings
Gupta et al. (2024)	IV	Mobile-phone-based agricultural advisory (KCC) in India	Access to phone-based extension services increased crop yields.
Cole and Fernando (2012)	RCT	Mobile-phone-based agricultural advisory in India	Randomized access to phone-based extension services increased farmers knowledge and input adoption, but impacts on yield were modest.
Fafchamps and Minten (2012)	DID / RCT hybrid	SMS-based market and weather information in Maharashtra	Random assignment of SMS information improved market participation but had limited yield effects; identification exploits village-level treatment timing.
Aker (2010)	IV / observational	Mobile phone rollout in Niger grain markets	Uses geographic variation in mobile coverage as an instrument; reduced price dispersion and improved welfarebenchmark non-experimental ICT study.
Goyal (2010)	Event study/ DID style	Computerized market information system in Central India	Instrumental variable approach exploiting phased rollout; improved farmer price realization and reduced trader margins.
Jensen (2007)	Natural experiment	Fishermens mobile phone adoption in Kerala	Exploits staggered introduction of mobile networks across coastal districts; showed efficiency and welfare gains through improved information flow.
Aker et al. (2012)	DID / RCT	Mobile learning and information platform in Niger	Randomized encouragement design; found gains in knowledge retention but heterogeneous yield responses.
Fabregas et al. (2019)	Review / meta-analysis	Review of digital agricultural extension interventions	Highlights predominance of RCT and DID designs; calls for complementary quasi-experimental and IV-based evidence at scale.
Blumenstock et al. (2016)	DID/IV	Mobile usage and airtime transfers in Rwanda	Uses lagged behavioral instruments for mobile adoption; demonstrates value of IV when self-selection drives digital engagement.
Toyama (2011)	Conceptual	Technology and human agency framework	Argues ICT acts as an amplifier of intentendogeneity of digital adoption requires empirical designs beyond simple mean comparison.
Deichmann et al. (2016)	Review	Digital agriculture in developing countries	Synthesizes RCT, DID, and observational studies; recommends hybrid identification strategies to address adoption endogeneity.

Table B.2: Spatial and Serial Correlation Test Results

A. Moran's I Spatial Autocorrelation Test			
Statistic	Value	Expected Value	Variance
Moran's I	-0.0715	-0.0005	0.0003
S.D.	-4.05		
p-value	1.00		
B. Serial Correlation Tests			
Test	Statistic	p-value	
Wooldridge (AR(1))	F = 6.14	0.013	
Breusch-Godfrey/Wooldridge	$\chi^2 = 4.47$	0.034	
Durbin-Watson	DW = 2.08	0.98	

Notes: Morans I assesses spatial autocorrelation in regression residuals. Serial correlation tests examine the presence of autocorrelation over time or panel units. Wooldridge, Durbin-Watson, and Breusch-Godfrey/Wooldridge refer to panel serial correlation test variants.

Table B.3: Oster δ-Robustness Analysis for Omitted Variable Bias

	δ^*	OLS (β_{OLS})	IV (β_{IV})	R^2_{OLS}	R^2_{max}
% Queries on Pest$_{(ds)ry}$	1.56	0.0016	-0.0291	0.819	1.000
% Queries on Weather$_{(ds)ry}$	0.27	0.0068	0.0077	0.819	1.000

Notes: δ^* is the implied ratio of selection on unobservables to observables needed to shift the OLS coefficient to the IV estimate, following Oster (2019). R^2_{OLS} refers to the controlled specification with fixed effects. Interpretation: $\delta^* > 1$ indicates strong robustness; $\delta^* < 1$ indicates sensitivity.

Table B.4: Decomposition of IV–OLS Coefficient Gap for Treatment Variables

	OLS	IV	IV–OLS Gap	Δ_{CW}	Δ_{TW}	Δ_{ME}
% Pest	0.0017^{**}	-0.0291^{**}	-0.0307^{**}	0.0004	-0.0021^{**}	-0.0290^{**}
Queries$_{(ds)ry}$	(0.0008)	(0.014)	(0.014)	(0.0003)	(0.0009)	(0.014)
% Weather	0.0067^{***}	0.0077^{*}	0.0010^{*}	-0.0002	-0.0063^{***}	0.0075^{*}
Queries$_{(ds)ry}$	(0.0006)	(0.005)	(0.005)	(0.0002)	(0.0006)	(0.005)

Notes: Robust standard errors in parentheses following bootstrap simulation for IV-OLS gap, covariate-weight, treatment weight and marginal-effect differences. Δ_{CW} = covariate-weight difference; Δ_{TW} = treatment-weight difference; Δ_{ME} = marginal-effect difference. OLS: Ordinary Least Square estimates, IV: Instrumental Variable estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.5: First-stage regressions on the percentage of seasonal queries at the district level on pest and weather for Paddy (treatment variables)

Instrumental Variables	Treatment variable: % Queries on pest		Treatment variable: % Queries on weather	
	(1)	(2)	(3)	(4)
Yield$_{(ds)r(y-1)}$ (2008-2018) (t ha^{-1})	1.043^{*}	1.041^{*}		
	(0.547)	(0.538)		
% Gross Irrigated area$_{(ds)y}$			-0.033^{***}	-0.020^{*}
			(0.012)	(0.012)
District $\times$ Season f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Additional controls	$\times$	✓	$\times$	✓
Observations	8,828	8,828	8,828	8,828
F-statistic	11.75	11.97	32.67	33.50

Note: The analysis involves district-level data from 10 states from 2009 to 2019. The KCC data in the analysis are related to the paddy crop for pest and weather queries. The table reports the first-stage regression of treatment variables (i.e., percentage seasonal queries on pests for the paddy crop and weather as outcome variables) on respective instrumental variables (previous years' yields and % gross irrigated area). Columns (1) and (3) report the estimates with baseline controls. Columns (2) and (4) include baseline and additional controls. The baseline controls include seasonal recorded maximum and minimum temperatures, and rainfall, deviations of temperature (minimum and maximum), and rainfall from historical averages, fertilizer consumption per gross cropped area, gross cropped area per total district area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. All regression equations include district $\times$ season-fixed effects. Standard errors are clustered at the district $\times$ season level reported in parentheses. The first-stage F-statistics values for the instruments' validity are reported in all columns. Subscripts for district d, State s, season r and year y. Significance level: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.6: Relation between IVs and Control Variables

	Dependent variables:	
	Yield$_{(ds)r(y-1)}$	% Gross Irrigated area$_{(ds)y}$
	(1)	(2)
Max Temp$_{(ds)ry}$(°C)	0.020^{**}	-0.187
	(0.010)	(0.120)
Min Temp$_{(ds)ry}$(°C)	0.001	0.086
	(0.015)	(0.224)
Rainfall$_{(ds)ry}$ (mm)	0.0001^{**}	0.0004
	(0.00005)	(0.001)
TD Max$_{(ds)ry}$	-0.070	-0.076
(dummy, 0 or 1)	(0.046)	(0.668)
TD Min$_{(ds)ry}$	0.046	-1.231^{**}
(dummy, 0 or 1)	(0.044)	(0.503)
RD$_{(ds)ry}$	0.019	-0.804
(dummy, 0 or 1)	(0.042)	(0.817)
Avg. landholding size$_{(ds)y}$	0.031	-2.521
(hectare)	(0.080)	(2.182)
NPK per GCA$_{(ds)y}$	0.00003	0.006^{*}
(kg/hectare)	(0.0002)	(0.003)
Gini coefficient$_{(ds)y}$	0.094	7.660
(land holding size)	(0.482)	(7.688)
% Gross Cropped Area (GCA)$_{(ds)y}$	0.001^{***}	-0.046^{**}
	(0.0005)	(0.020)
# Markets$_{(ds)y}$	-0.008	-0.856^{***}
	(0.008)	(0.147)
# ULB$_{(ds)}$ (i.e., municipalities)	4.976^{***}	105.629^{***}
	(1.244)	(9.049)
Road length per total	-0.007	0.099
dist. area$_{(ds)y}$ (km/hectare)	(0.005)	(0.121)
# Train stations per total	-298.896^{***}	-14381.070^{***}
dist. area$_{(ds)}$	(89.949)	$(1,000.426)$
% Rural population$_{(ds)y}$	0.003	-0.061
	(0.006)	(0.077)
District f.e.	✓	✓
Season f.e.	✓	✓
District × Season f.e.	✓	✓
Year f.e.	✓	✓
Observations	8,828	8,828
F-Statistic	28.38	57.74
R^2	0.82	0.88

Note: The table displays the outcomes of the interactions between instrumental and control variables to validate the exclusion criterion. If the instruments exhibit strong correlations with the observable variables (such as control variables), they are likely correlated with unobservable factors as well (Altonji et al., 2011). Columns (1) and (2) present these interactions with past years' yields (acting as IVs for queries related to paddy pests) and the percentage of gross irrigated area (serving as IVs for weather-related queries) as the dependent variables, respectively. Heteroskedasticity-robust standard errors clustered at the district × season level are reported in parentheses. Subscripts for district d, State s, season r and year y. Significance level: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.7: Testing for positive 2SLS weights

Instrumental Variables	Treatment Variables: % Queries on	
	Pest	Weather
	(1)	(2)
	Panel A: Continuous variables results	
$\text{Yield}_{(ds)r(y-1)}$	1.041^{*}	
	(0.538)	
weight	0.0037	
p-value: positive weights	0.487	
p-value: negative weights	0.513	
$\%$ Gross Irrigated area$_{(ds)y}$		-0.020^{*}
		(0.012)
weight		0.0938
p-value: positive weights		1.000
p-value: negative weights		0.000
	Panel B: Binary variables results	
$\text{Yield}_{(ds)r(y-1)}$	0.079^{***}	
	(0.019)	
weight	0.0338	
p-value: positive weights	0.995	
p-value: negative weights	0.005	
$\%$ Gross Irrigated area$_{(ds)y}$		-0.320
		(0.021)
weight		0.0449
p-value: positive weights		1.000
p-value: negative weights		0.000
Observations	8,828	8,828
District $\times$ Season f.e.	✓	✓
Year f.e.	✓	✓
Baseline controls	✓	✓
Additional controls	✓	✓

Note: The analysis involves district-level data from 10 states from 2009 to 2019. The KCC data in the analysis are related to the paddy crop for pest and weather queries. The table reports the first-stage regression of treatment variables (i.e., percentage seasonal queries on pests for the paddy crop and weather) on respective instrumental variables (previous years' yields and % gross irrigated area). Panel A represents treatment and continuous variables as continuous variables similar to results under (Table B.5). Panel B represents treatment and instrumental variables with as binary variables (i.e., taking a value 1 if the variables values are greater than their respective means). All regression equations include district $\times$ season-fixed effects, baseline controls, and additional controls. The baseline controls include; seasonal recorded maximum and minimum temperatures, and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include; the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Standard errors are clustered at the district $\times$ season level reported in parenthesis. The initial p-value derives from testing the null hypothesis asserting all 2SLS weights are positive, while the subsequent p-value evaluates the hypothesis that at least one weight is negative for the respective instruments across each panel. Subscripts for district d, State s, season r and year y. Significance level: *p<0.1; **p<0.05; ***p<0.01.

Table B.8: Regression results from zero-first-stage test for Paddy

Instrumental Variables	Treatment variables: % Queries on -		Outcome variable	
	Pest	Weather	$\text{Yield}_{(ds)ry}$	
	(1)	(2)	(3)	(4)
	Reduced form (full sample)			
$\text{Yield}_{(ds)r(y-1)}$	1.041^{*}		-0.00002	
	(0.538)		(0.031)	
% Gross Irrigated $\text{area}_{(ds)y}$		-0.020^{*}		0.002^{**}
		(0.012)		(0.001)
F-stat on instruments	11.97	33.5	28.05	29.17
Observations	8,828	8,828	8,828	8,828
	Direct effect (zero-first-stage sample group)			
$\text{Yield}_{(ds)r(y-1)}$	0.500		-0.010	
	(0.812)		(0.050)	
% Gross Irrigated $\text{area}_{(ds)y}$		-0.004		0.001
		(0.026)		(0.002)
F-stat on instruments	14.31	46.29	21.04	88.57
Observations	4,130	2,652	4,130	2,652
% of Total observations	46.78%	30.04%	46.78%	30.04%
Selected States	BH, OD, GJ, HY, PB	BH, MH, PB	BH, OD, GJ, HY, PB	BH, MH, PB
	Direct effect (remaining sample group)			
$\text{Yield}_{(ds)r(y-1)}$	1.254^{*}		-0.001	
	(0.674)		(0.038)	
% Gross Irrigated $\text{area}_{(ds)y}$		0.035^{**}		0.002^{**}
		(0.015)		(0.001)
F-stat on instruments	10.11	35.17	30.81	21.82
Observations	4,698	6,176	4,698	6,176
% of Total observations	53.22%	69.96%	53.22%	69.96%
Remaining States	MH, MP, AP, TG, TN	HY, MP, GJ OD, AP, TG, TN	MH, MP, AP, TG, TN	HY, MP, GJ OD, AP, TG, TN

Note: All specifications include the unreported constant term, baseline control, other control, and district $\times$ season-fixed effects. Heteroskedasticity-robust standard errors clustered at the district $\times$ season level are reported in parentheses. Subscripts for district d, State s, season r and year y. Abbreviations: **BH-Bihar, OD-Odisha, GJ-Gujarat, MH-Maharashtra, PB-Punjab, HY-Haryana**, MP-Madhya Pradesh, AP-Andhra Pradesh, TG-Telangana and TN-Tamil Nadu. Significance level: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.9: Regression results on the relevance of instruments with variables other than the treatment variable for Paddy crop

Instrumental Variables	**Dependent Variables:**			
	NPK/GCA	**Rainfall**	**Price Shock**	
	(1)	(2)	(3)	(4)
Yield$_{(ds)r(y-1)}$	0.419		0.003	
	(1.122)		(0.005)	
% Gross Irrigated Area$_{(ds)y}$		0.027		-0.0003
		(0.136)		(0.0002)
District $\times$ Season f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓
R^2	0.87	0.92	0.17	0.14
F-stat on instruments	45.62	94.40	1.38	1.34
Observations	8,828	8,828	8,828	8,828

Note: The table presents results testing the relevance of our instruments (previous year's yields for % seasonal pest queries and % gross irrigated area for % seasonal weather queries) with variables other than the treatment variables (fertilizer consumption (NPK) per GCA, rainfall and price shocks). The dependent variables are shown at the top. All regressions include an unreported constant term, baseline controls, additional controls, and district, season, and year fixed effects. Heteroskedasticity-robust standard errors clustered at the district $\times$ season level are reported in parentheses. Subscripts for district d, state s, season r, and year y. Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.10: Baseline regressions for Paddy

Variables	Dependent variable: Paddy Yield (t ha^{-1})			
	(1)	(2)	(3)	(4)
% Queries on Pest$_{(ds)ry}$	0.002***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
% Queries on Weather$_{(ds)ry}$	0.001	0.001	0.002*	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Max Temp$_{(ds)ry}$(°C)		0.020**	0.045***	0.038***
		(0.008)	(0.010)	(0.010)
Min Temp$_{(ds)ry}$(°C)		−0.061***	−0.060***	−0.066***
		(0.016)	(0.018)	(0.017)
Rainfall$_{(ds)ry}$ (mm)		−0.00002	0.0001**	0.0001***
		(0.00004)	(0.0001)	(0.0001)
TD Max$_{(ds)ry}$		−0.011	−0.049	−0.037
(dummy, 0 or 1)		(0.049)	(0.047)	(0.044)
TD Min$_{(ds)ry}$		0.012	0.035	0.060
(dummy, 0 or 1)		(0.051)	(0.046)	(0.046)
RD$_{(ds)ry}$		−0.015	−0.020	0.011
(dummy, 0 or 1)		(0.058)	(0.059)	(0.056)
Avg. landholding size$_{(ds)y}$		−0.0002	0.014	0.012
(hectare)		(0.081)	(0.082)	(0.077)
NPK per GCA$_{(ds)y}$		0.0004	0.0003	0.0004*
(kg/hectare)		(0.0003)	(0.0002)	(0.0002)
Gini coefficient$_{(ds)y}$		−0.167	−0.071	−0.709
(land holding size)		(0.727)	(0.731)	(0.709)
% Gross Cropped Area (GCA)$_{(ds)y}$		−0.0002	−0.0003	−0.0005
		(0.0005)	(0.0005)	(0.0005)
# Markets$_{(ds)y}$		−0.004	0.003	0.003
		(0.008)	(0.007)	(0.007)
# ULB$_{(ds)}$ (i.e., municipalities)		4.508***	3.462**	2.884*
		(1.074)	(1.379)	(1.275)
Road length per total		−0.001	−0.001	0.0003
dist. area$_{(ds)y}$ (km/hectare)		(0.005)	(0.005)	(0.005)
# Train stations per total		−283.033***	−206.852**	−173.490*
dist. area$_{(ds)}$		(77.097)	(99.149)	(91.824)
% Rural population$_{(ds)y}$		−0.007	−0.007	−0.010
		(0.009)	(0.009)	(0.009)
Yield$_{(ds)r(y-1)}$				−0.002
				(0.028)
District f.e.	✓	✓	✓	✓
Season f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓
District × Season f.e.	×	×	✓	✓
Observations	8,828	8,828	8,828	8,828
F-Statistic	36.39	35.19	29.29	29.15
R^2	0.80	0.80	0.83	0.82

Note: The table presents the results of the baseline regressions. Column (1) excludes the baseline and other control variables, whereas columns (2) and (3) include the baseline and other control variables. All specifications include the unreported constant term. Heteroskedasticity-robust standard errors clustered at the district × season level are reported in parentheses. Subscripts for district d, State s, season r and year y. Significance level: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table B.11: Falsification, Placebo and Non-linearity checks: Paddy yields and farmer's queries

Treatment Variables	Dependent Variables (t ha^{-1})				
	Wheat yields	Pulses yields	Paddy yields		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Panel A: Falsification check - Alternate Dependent Variables				
% Queries on Pest$_{(ds)ry}$	−0.079	0.020			
	(0.174)	(0.034)			
% Queries on Weather$_{(ds)ry}$	−0.238	0.018			
	(0.599)	(0.021)			
	Panel B: Placebo check - Alternate Treatment Variables				
% Queries on Pest$_{(ds)r(y+1)}$			−0.109		
			(0.076)		
% Queries on Weather$_{(ds)r(y+1)}$			0.003		
			(0.012)		
	Panel C: Nonlinearity Check - Cubic B-Splines Model				
LR: % Queries on Pest$_{(ds)ry}$				20.554	
				(236.320)	
MR: % Queries on Pest$_{(ds)ry}$				−7.387	
				(74.891)	
HR: % Queries on Pest$_{(ds)ry}$				22.192	
				(241.486)	
LR: % Queries on Weather$_{(ds)ry}$				−22.653	
				(215.315)	
MR: % Queries on Weather$_{(ds)ry}$				17.527	
				(144.946)	
HR: % Queries on Weather$_{(ds)ry}$				−0.253	
				(144.946)	
	Panel D: Nonlinearity Check - Quadratic Model				
% Queries on Pest$_{(ds)ry}$					−0.081
					(0.155)
% Queries on Pest$^2_{(ds)ry}$					0.001
					(0.003)
% Queries on Weather$_{(ds)ry}$					0.031
					(0.064)
% Queries on Weather$^2_{(ds)ry}$					−0.001
					(0.002)
Observations	1,835	2,434	8,828	8,828	8,828
Wu-Hausmann statistic	19.25***	1.085	8.093***	9.985***	NA
p-value	(5.59×10^{-9})	(0.338)	(0.0003)	(0.0001)	NA
District × Season f.e.	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓

Note: The table reports 2SLS-IV estimates on the effect of percentage seasonal queries on pest and weather at the district level on paddy yields. All specifications include the instrumental variables; the previous year's paddy yield for pest queries and the % gross irrigated area for weather queries. Panel A, B, C and D represents alternate dependent variables (wheat and pulses), alternate treatments (following years pest and weather related queries), cubic B-splines estimating threshold effect and quadratic models, respectively . All regression equations include district × season-fixed effects, baseline controls, and additional controls. The baseline controls include; seasonal recorded maximum and minimum temperatures, and rainfall, deviations of temperatures (minimum and maximum) and rainfall from historical averages, total fertilizer consumption per gross cropped area, percentage gross cropped area, Gini coefficient (agricultural land holding size), average agricultural land holding and percentage rural population. The additional controls include; the number of markets, urban local bodies (i.e., municipalities), and road length and train stations per total district area. Standard errors are clustered at the district × season level reported in parenthesis. Subscripts for district d, State s, season r and year y. LR: Low range, MR: Medium range, HR: High range. Significance level: *p<0.1; **p<0.05; ***p<0.01.

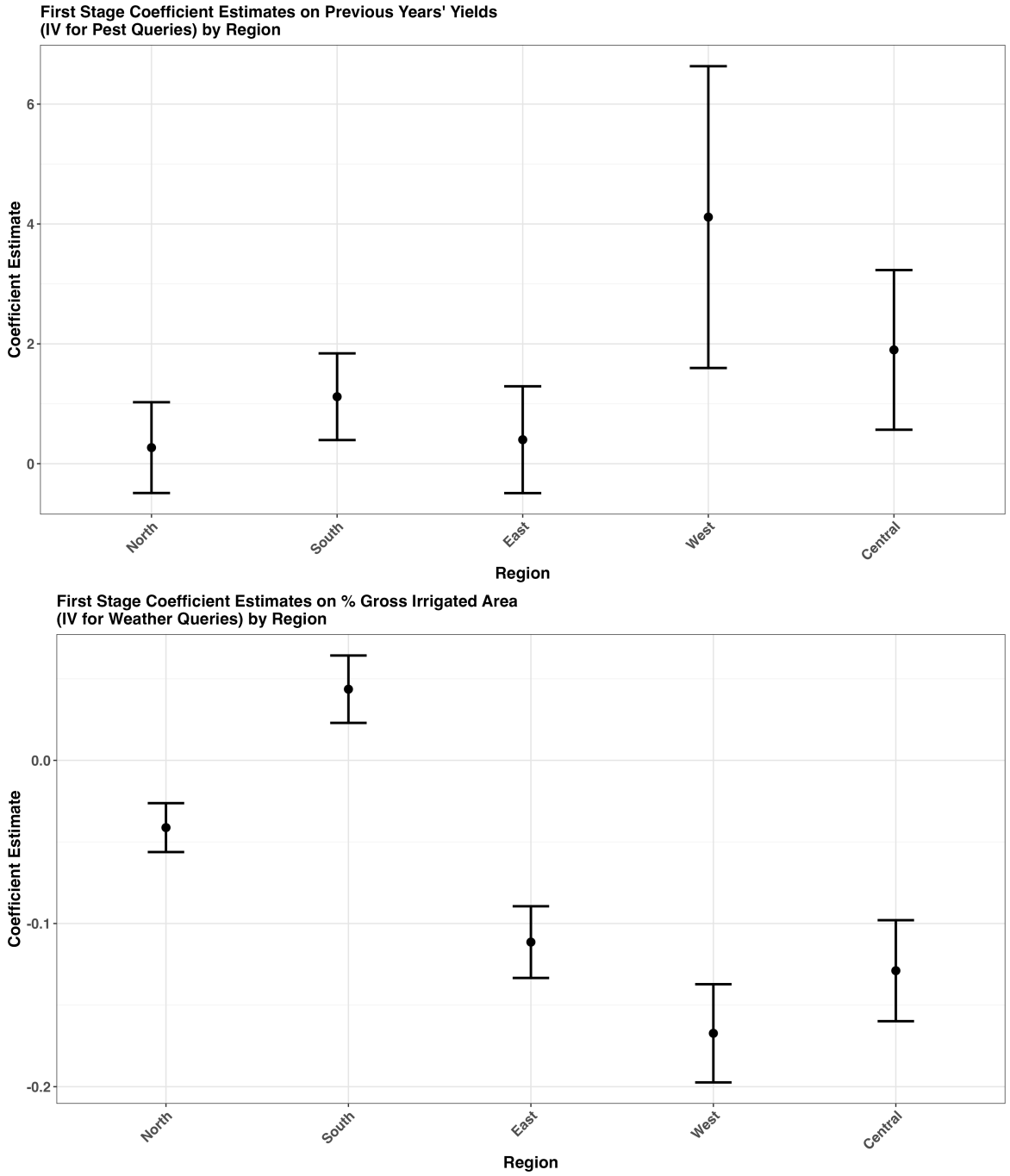


Figure B.1: Coefficient estimates of the Instrumental Variables (Previous years' paddy yields and % Gross Irrigated area) based on first-stage regressions by Region

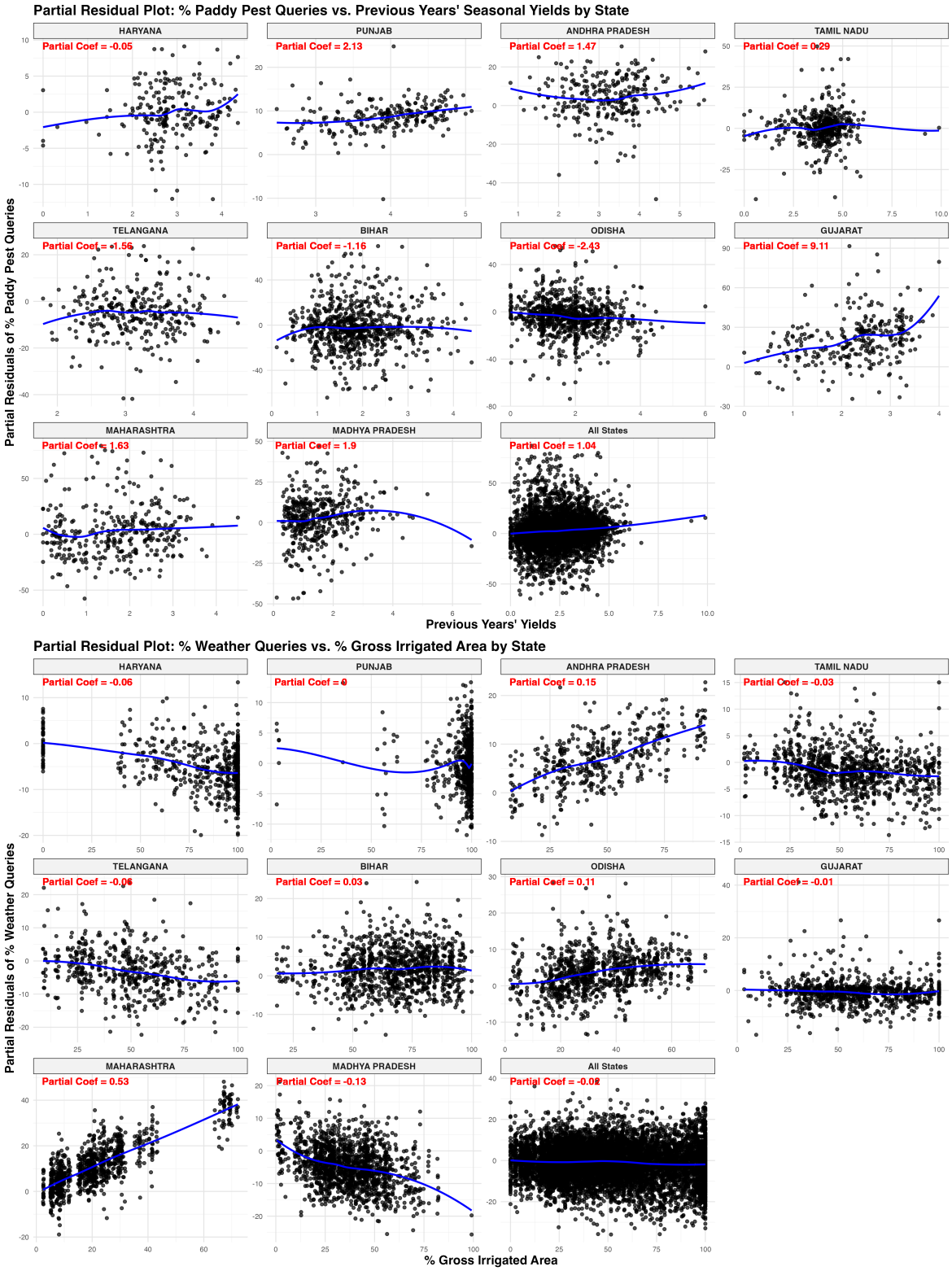


Figure B.2: Coefficient estimates and residual plots of the Instrumental Variables (Previous years' paddy yields and % Gross Irrigated area) based on first-stage regressions by State

C Additional evidence for selecting the States for zero-samples

Indian crop production follows a seasonal cropping pattern aligned with monsoon rains (known as the *Kharif* season), summer, and winter (the *Rabi* season). The monthly distribution of these cropping seasons varies according to agroclimatic regions, geographic regions, and States. The paddy yield data are divided into two consecutive five-year periods for comparative analysis. The first period (2010-2014), with the start year matching the 2010-11 agricultural census. The second period covers 2015-2019, beginning with the 2015-16 agricultural census year. Comparing district-level yields between these two periods allows one to assess yield changes over time, indicating hot/coldspots.

The methodology for data analysis involves the comparison of district-level seasonal yield variations with the all-India seasonal yield averages for paddy, across two 5-year periods and three cropping seasons as described in equation (C.1). The mean district-level yield variances for each period are visualized on maps to pinpoint regions with either higher or lower crop yields. Additionally, we identify crop yield hotspots, which denote significant positive district-level seasonal crop yield differences between the two periods. These hotspots encompass districts with crop acreage exceeding the national average in respective periods or districts ranking above the 80th percentile in yield differences based on seasons and period. Conversely, districts exhibiting statistically significant negative differences in seasonal crop yields compared to the national averages for the two periods are termed coldspots. The statistical significance of the yield differences is determined using a two-sample Student's t-test, following equation (C.2).

The estimation equation for the average seasonal crop yield differences is as follows.

$$\overline{YieldDiff}_{(ds)kp_i} = \frac{1}{5}\sum_{y_{p_i}=1}^{5}(Yield_{(ds)ky_{p_i}} - \frac{1}{5\cdot n}\sum_{y_{p_i}=1}^{5}\sum_{(ds)=1}^{n} Yield_{(ds)ky_{p_i}}) \tag{C.1}$$

where, $\overline{YieldDiff}$: Average yield difference, (ds): District-State, n: Total number of districts (732), k: Seasons, p_i: i^{th} period (i =1,2) and y: Year sequence in a period (p_i).

Student's two-sample t-test: The test to compare district-level yield differences between periods p_1 (2010-2014) and p_2 (2015-2019) based on seasonal crop yield averages. The standard t-test formulae are illustrated in our context as follows:

$$t = \frac{\overline{YieldDiff}_{p_1} - \overline{YieldDiff}_{p_2}}{\sqrt{s^2(\frac{1}{m_1} + \frac{1}{m_2})}} \tag{C.2}$$

$$s^2 = \frac{\sum_{i=1}^{m_1}(YieldDiff_i - \overline{YieldDiff}_i)^2 + \sum_{j=1}^{m_2}(YieldDiff_j - \overline{YieldDiff}_j)^2}{m_1 + m_2 - 2}$$

where, t : Student's t-statistic, s : standard deviation, m_i : number of observations for crop yields in respective districts and corresponding period (p_i).

The seasonal paddy yield differences and hotspot maps are illustrated in Figure C.1 for the monsoon season, Figure C.2 for the winter season, and Figure C.3 for the summer season. The States of Gujarat and Maharastra cultivate paddy exclusively during the monsoon season and are categorized as low-performing zone. On the other hand, the districts of Bihar and Odisha grow paddy in all three seasons and remain consistently in the category of underperforming States.

We also plotted the average percentage of the gross irrigated area in all districts of India (Figure C.4). The States of Maharashtra and Odisha are categorized as having limited availability or use of irrigation facilities. On the other hand, Bihar's irrigation utilization surpasses the national average but has yet to reach the potential yield for paddy.

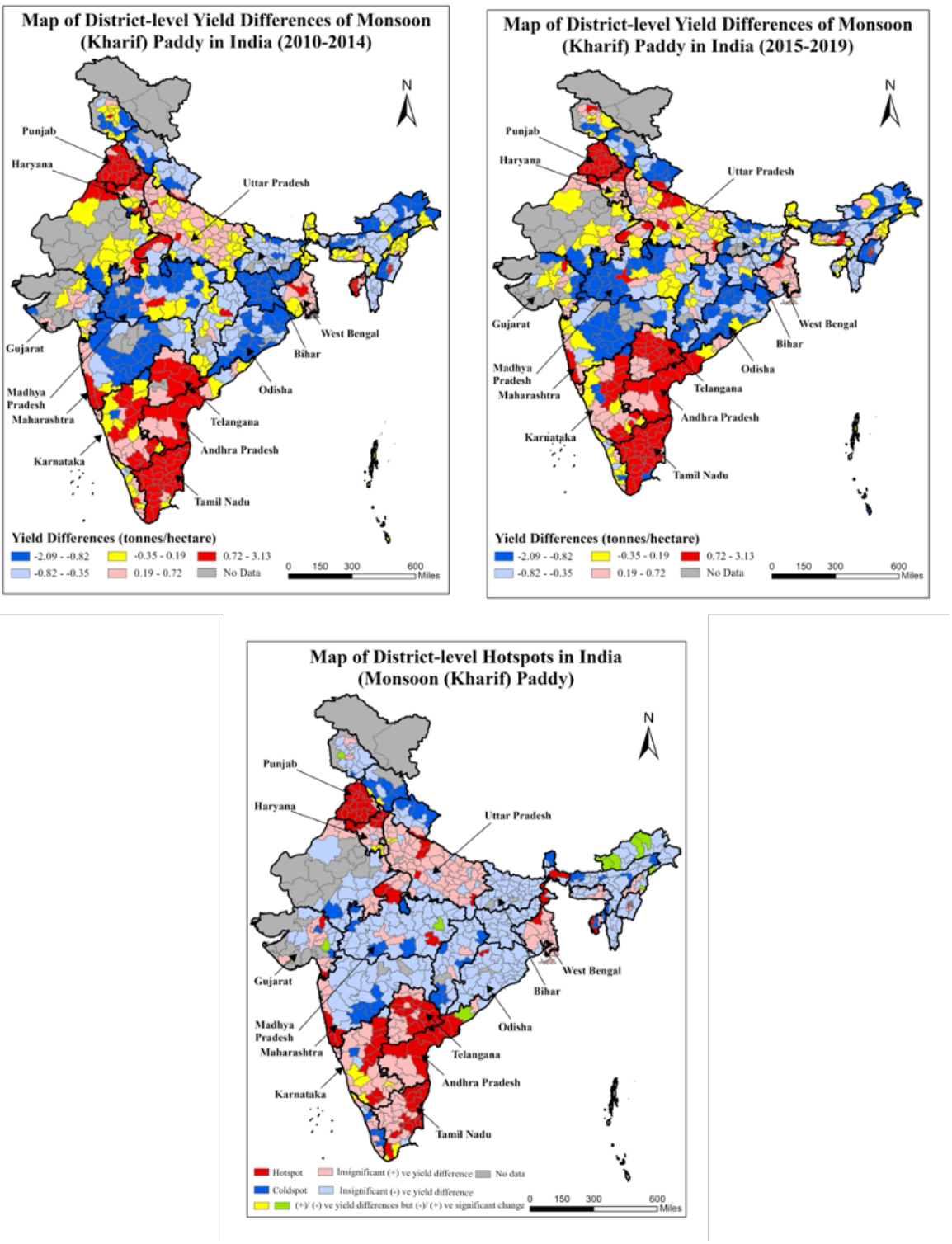


Figure C.1: Maps of district-level yield differences for the two-periods and hotspots for monsoon (*Kharif*) paddy in India

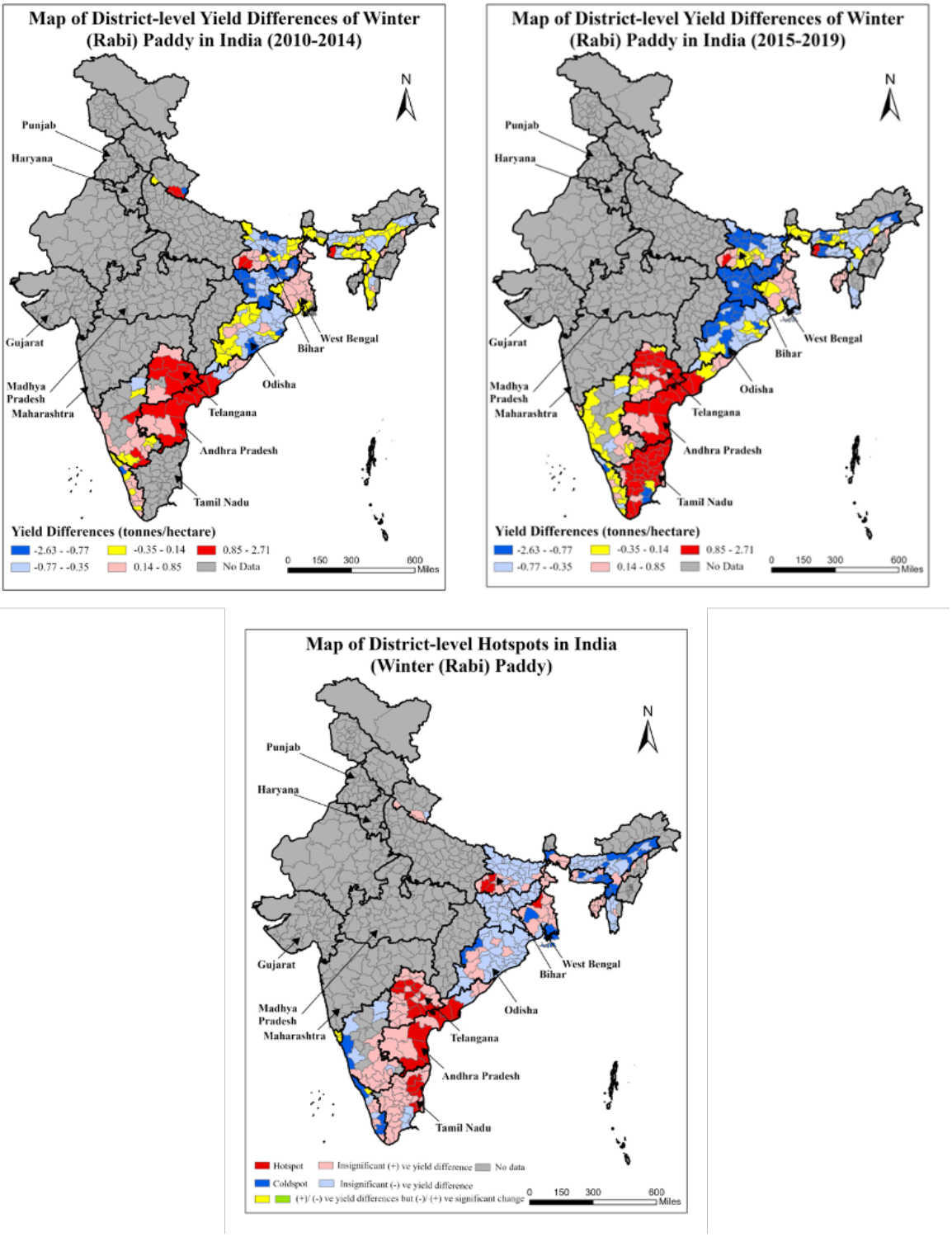


Figure C.2: Maps of district-level yield differences for the two-periods and hotspots for winter (*Rabi*) paddy in India

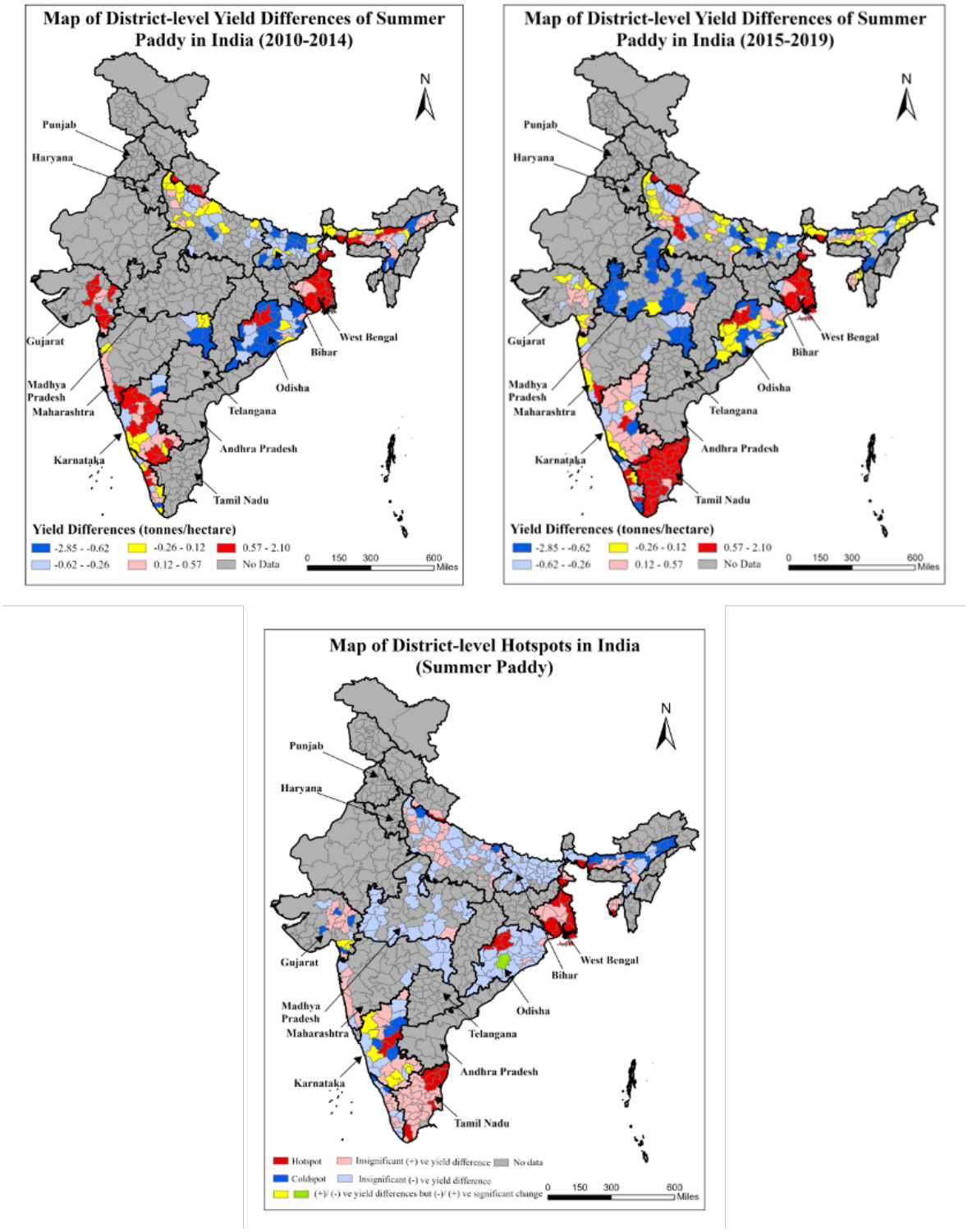


Figure C.3: Maps of district-level yield differences for the two-periods and hotspots for summer paddy in India

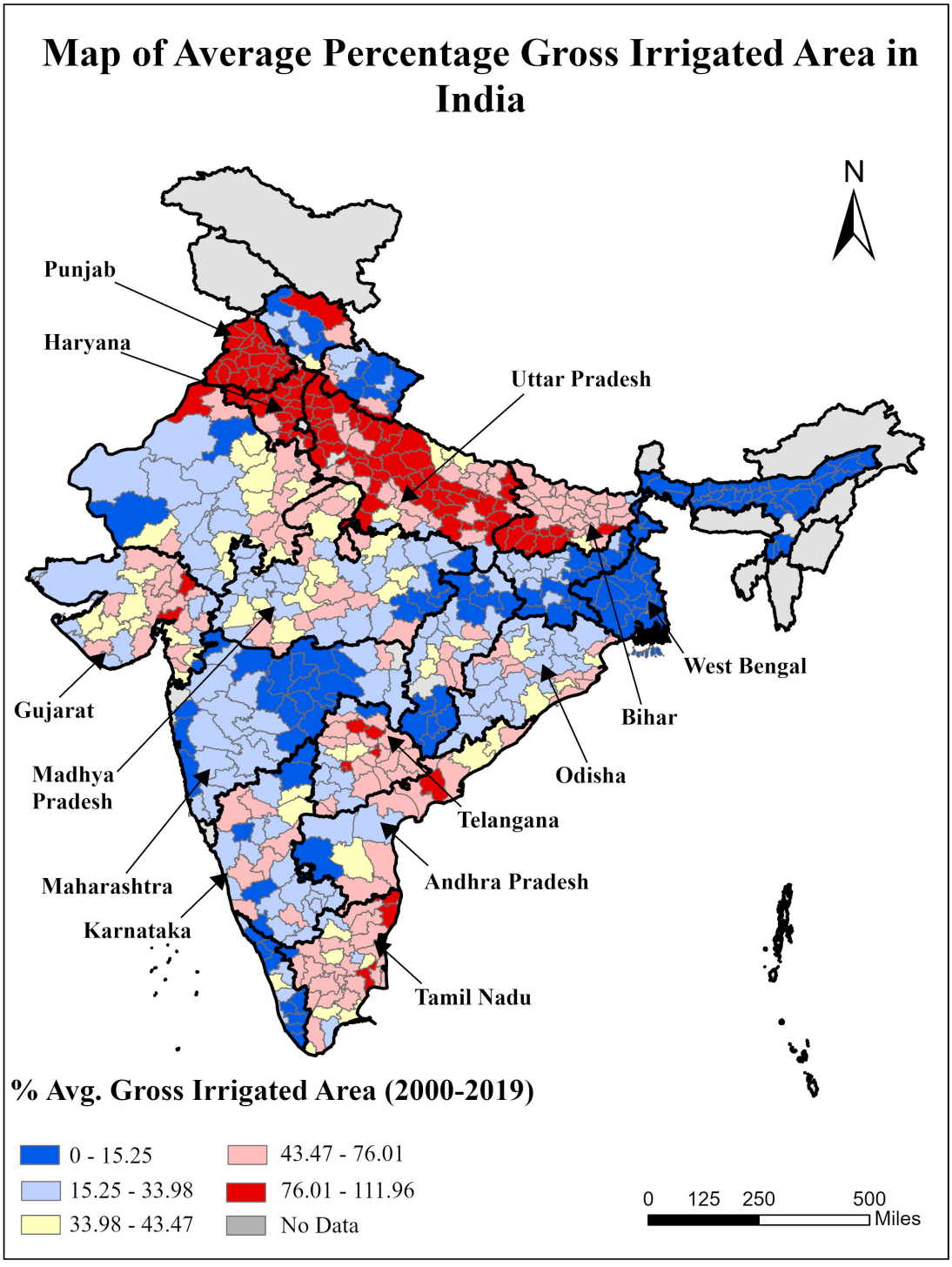


Figure C.4: Maps of district-level average percentage gross irrigated area in India