Language Barriers, Technology Adoption and Productivity: Evidence from Agriculture in India*

Apoorv Gupta† Jacopo Ponticelli‡ Andrea Tesei§

This draft: April 2024

Abstract

We study the effect of language barriers on the ability of farmers to access information about agricultural technologies in rural areas of India. We use the introduction of government-sponsored call centers (Kisan Call Centers) which offer agricultural advice in the official language of each Indian state. For identification, we compare geographically contiguous areas that sit across state borders, and exploit differences in the language spoken by farmers and call center advisors. We document that language barriers limit the adoption of modern agricultural technologies – such as high-yielding variety seeds – and negatively affect crop yields.

Keywords: Official Languages of India, HYV Seeds, Kisan Call Centers.
JEL Classification: O10, Q16, Z13

* We received valuable comments from Siwan Anderson, Shawn Cole, Diego Comin, Nicola Gennaioli, Doug Gollin, Sergei Guriev, Sabrina Howell, Dean Karlan, Marco Manacorda, Pepita Miquel, Imran Rasul, Gabriella Santangelo, Chris Udry and seminar participants at Columbia University, Berkeley Haas, UBC, University of Zurich, University of Bonn, NYU Stern, Bocconi University, Queen Mary University, Lancaster University, University of Maryland, Northwestern University, CUNY, Diego Portales University, Universidad de Los Andes, Universidad Javeriana, University of Geneva, GATE-Lyon, BGSE Summer Forum, CAF, DFID UK, Federal Communications Commission, AFES 2022, STEG Workshop 2021, European Economic Association 2019, IPA 2019 Researcher Gathering and American Economic Association 2021. The project was developed while Andrea Tesei was visiting the Ford Motor Company Center for Global Citizenship at the Kellogg School of Management, Northwestern University, whose hospitality is gratefully acknowledged. Anoushka Nalwa, Jora Li, Pierre Jaffard, Gursharan Bhue and Mark He provided excellent research assistance. An earlier version of this paper circulated under the title "Access to Information, Technology Adoption and Productivity: Large-scale Evidence from Agriculture in India".

† Dartmouth College; contact: apoorv.gupta@dartmouth.edu.

‡ Northwestern University - Kellogg School of Management, NBER & CEPR; contact: jacopo.ponticelli@kellogg.northwestern.edu.

§ Queen Mary University of London, CEPR, CEP (LSE) & CESifo; contact: a.tesei@qmul.ac.uk.

I Introduction

Language differences between individuals impose higher transaction costs for the acquisition of information and may result in slower learning about new technologies and economic opportunities. The issue is particularly relevant in areas characterized by high levels of linguistic fragmentation and far from the technological frontier, such as agricultural regions of developing countries (Eberhard et al., 2022). Previous studies have shown that the modernization of agriculture in these areas is often limited by farmers' imperfect information on new technologies (e.g., Foster and Rosenzweig, 1995; Conley and Udry, 2010). However, we still have scarce direct empirical evidence on how language barriers affect the dissemination of such information.

In this paper, we aim to fill this gap by studying the impact of language barriers on the adoption of modern technologies in agriculture. We investigate this question in the context of India, which is well suited for a number of reasons. India has 22 languages officially recognized in its Constitution and 99 non-officially recognized languages, each spoken by at least 10,000 people. This implies a level of language fragmentation comparable to that of Sub-Saharan Africa (Easterly and Levine, 1997). In addition, India has an under-served demand for information among rural farmers. Still in 2003, 60% of Indian farmers reported not having access to any source of information on modern agricultural technologies (National Sample Survey, 2005). Finally, the Indian context offers a natural experiment that allows us to make progress in disentangling the role of language barriers from other characteristics that also tend to differ across groups speaking different languages and that may explain different rates of technology adoption (Ginsburgh and Weber, 2020).

To isolate the role of language barriers, we compare speakers of official Indian languages that have different access to a new phone-based information platform for agricultural advice called Kisan Call Centers (KCC). A key feature of KCC is that the service is only offered in the official language of each Indian state. This generates differences in potential access to the service across geographically contiguous areas (10×10 km cells) that sit across state borders, whenever the official language of a state does not match the official language spoken by the underlying population. As such, our identification strategy exploits differences in the *composition* of official language speakers across contiguous areas, conditional on the share of individuals speaking any of the official languages of India. These contiguous areas are comparable across a large set of socio-economic and ethnic characteristics. Using event-study analysis, we show that these areas follow similar trends in technology adoption and productivity prior to the introduction of KCC.

We combine three main data sources. First, data on the location and content of all phone calls made by farmers to KCC between 2006 and 2017. This allows us to observe farmers' questions about specific agricultural technologies and the answers they receive from agronomists. Second, we use data on the adoption of agricultural technologies from the Agricultural Input Survey of India, which was carried out at 5-year intervals between 2002 and 2017. This survey includes information on farmers' adoption of agricultural inputs including high-yielding variety (HYV) seeds, chemical fertilizers and artificial irrigation systems. HYV seeds are commercially developed to increase crop yields and are one of the most important innovations in modern agriculture (Evenson and Gollin, 2003). Chemical fertilizers and reliable irrigation systems are key complementary inputs to maximize HYV potential. Third, to measure farmers' yields, we use data on area farmed and quantity of crops produced from ICRISAT, as well as changes in vegetation indices estimated from MODIS satellite images. The combination of these dataset allows us to map farmers' calls about agricultural technologies with their actual adoption, and then study their impact on agricultural productivity.

We document three key findings. First, areas with higher language barriers between farmers and agricultural advisors experience a significantly lower increase in the number of calls to KCC following the launch of the program. Our estimates indicate that areas with one standard deviation higher language barriers experienced 0.9 less calls to KCC per 100 farmers per year (37% of the sample mean) after the introduction of the program. This is consistent with language differences significantly affecting farmers' ability to access agriculture-related information. Second, we find that areas with higher language barriers between farmers and KCC advisors experience a significantly lower increase in the adoption of certain agricultural technologies, including HYV seeds (1.3% lower adoption for a standard deviation higher language barriers), fertilizers and artificial irrigation. These effects materialize within five years from the introduction of KCC and persist in the long run. Third, we find a negative effect of language barriers on agricultural productivity (0.6% lower crop yields for a standard deviation higher language barriers), although these effects tend to be less precisely estimated when we use satellite-based measures of agricultural productivity.

Our findings speak to three streams of the literature. First, the literature studying the effects of language diversity on economic and political outcomes. Previous studies have shown that greater linguistic distance between countries is associated with less bilateral trust and trade (Guiso et al., 2009; Melitz, 2008), less international migration (Adsera and Pytlikova, 2015) and larger cross-country differences in per capita income (Spolaore and Wacziarg, 2009). Within countries, greater language fragmentation is associated with less redistribution and public good provision (Alesina and Glaeser, 2004; Desmet et al., 2012; Ban et al., 2012), greater risk of conflict (Fearon and Laitin, 2003) and lower economic growth (Easterly and Levine, 1997). In the context of India, Fenske and Kala (2021) show how linguistic distance between regions affect their degree of market integration, and Jain (2017) documents how the mismatch between official languages and local languages affect educational attainment.[1] Our contribution to this literature is to focus on information acquisition about technology as a specific channel through which language differences can directly influence economic development, and present micro-based empirical evidence consistent with its effects. Our findings also relate to recent work on the effect of language differences on information diffusion within firms (Debaere et al., 2013; Guillouet et al., 2021). Compared to these studies, we focus on a setting – rural agricultural communities in developing countries – where language barriers are likely to be stronger, and focus on different outcomes such as technology adoption and productivity.

Second, we speak to the literature investigating the role of modern agricultural technologies – such as HYV seeds – in the process of development. This literature has studied several potential frictions to the adoption of modern technologies by farmers, including credit constraints, missing insurance markets, and lack of access to high-quality inputs (see Bridle et al. (2020) and Suri and Udry (2022) for recent reviews). Among these frictions, the lack of information on new technologies or how to use them has received extensive attention. This literature includes work grounded on learning models of new technologies based on farmers' own experience or the experience of others in their social network (Beaman et al., 2021; Conley and Udry, 2010; Foster and Rosenzweig, 1995; Hanna et al., 2014; Munshi, 2004). Still, within the literature on information frictions, there is scarce empirical evidence on the role of linguistic fragmentation. Our contribution to this literature is to focus on the role of language barriers between farmers and agricultural advisors as a friction to information diffusion.

Finally, the paper is related to the literature using randomized controlled trials to evaluate the impact of agricultural extension services (Aker et al., 2016; Fabregas et al., 2019). Previous research has highlighted the poor performance of traditional face-to-face programs that cannot provide timely and personalized information to farmers (Anderson and Feder, 2004; Duflo et al., 2011). KCC gives farmers access to customized and timely information throughout the agricultural cycle. We document that the availability of such information facilitates technology adoption and improves productivity. Existing evidence on the impact of mobile based intervention programs has documented significant effects on farmers' input choices but limited impact on their productivity. Casaburi et al. (2019) show that text messages containing agricultural advice have positive effects on the yields of small sugarcane farmers in Kenya, but the increase dissipates over time. Cole and Fernando (2020) randomize access to a hot line for agricultural advice to farmers in Gujarat, finding a significant impact on agricultural practices, but no systematic impact on yields. Fafchamps and Minten (2012) study the impact of a text message-based agricultural information system providing market and weather information to Indian farmers and find non-significant effects on cultivation practices or productivity.

[1] On the relationship between ethnic diversity and access to information in India see also Armand et al. (2022).

II Institutional Background and Empirical Strategy

In the mid-2000s, the Indian Ministry of Agriculture introduced the Kisan Call Centers (KCC) initiative, a set of call centers offering agricultural advice to farmers. Farmers can contact these call centers free of charge via landline or mobile phones. Calls are answered by trained agronomists, who address farmers' questions with advise that is specific to the agro-climatic characteristics of the area where the farmer is located. The Ministry of Agriculture opened 21 such call centers, which answer calls from all Indian states. As shown in Figure C.1(a), KCC received less than 1,000 calls per year in the first years after its introduction. The number of calls increased substantially starting in 2009, with between 500,000 and 1 million calls per year between 2009 and 2012 thanks to a large advertising campaign by the Ministry of Agriculture. The annual number of calls increased further between 2013 and 2017, with more than 4 million calls per year starting in 2015.

We use data on the universe of calls received by KCC between 2006 and 2017. The data reports call-level information on the question asked by the farmer, a brief description of the answer provided, and the time and location (subdistrict) from which the call was originated. As an illustration, Figure C.1(b) and (c) report the breakdown of calls by calendar month and topic for farmers asking questions related to the cultivation of rice and wheat – the two largest crops in India by area farmed. Rice farmers mostly ask questions about seeds in May and June – at the beginning of the *kharif* season. During the growing season, in July and August, calls about fertilizers increase. Finally, as crops fully grow and harvesting season approaches, most calls are about pests. Similar patterns can be observed for wheat, which is mainly farmed during the *rabi* season, in which crops are grown between October and November and harvested between December and the Spring months.

Our empirical analysis exploits a key institutional feature of KCC, namely that the service is only offered in the official language of the Indian state where the phone number of the caller is registered. This implies that only farmers speaking the official language of their state are able to ask questions and understand the answers provided by KCC agronomists. Figure I(a) reports the distribution of official languages by state. Since the State Reorganization Act of 1956 drew state borders along linguistic lines, the diffusion of Indian languages is relatively homogeneous within states (Jain, 2017). However, as shown in Figure I(b), the overlap between linguistic and administrative boundaries is not perfect and the share of people whose first language is an official Indian language other than the one of the state where they live tends to increase near state borders. This generates differences in potential access to the service between geographically contiguous areas located across state borders, which we exploit in the empirical analysis.

The geographical unit of observation in our empirical analysis is a 10×10 km cell. We use cells to match information from the main datasets used in the empirical analysis, which come at different levels of geographical aggregation.[2] In all our specifications, we focus on cells located within 50 km from state borders – excluding cells intersected by state borders – and show robustness to different thresholds in Section IV.

Figure I(c) reports the spatial distribution of the cells in the regression sample. Summary statistics are reported in Table I. Cells in our sample tend to be rural and specialized in agriculture. In the average cell, 91.6% of individuals speak one of the official languages of India, but around 13% of them do not speak the official language of the state in which they live. On average, we observe 2.5 calls to KCC per 100 farmers per year. We label calls about technology as those in which farmers ask questions about seed varieties, pesticides, fertilizers and irrigation. These categories account for 43% of all calls.[3]

One important reason for focusing on cells that are geographically close but on opposite sides of state borders is that state borders tend to generate discontinuities in the share of non-state language speakers among speakers of official languages. Figure I(d) plots the distribution of such gaps in the share of non-state language speakers between cells across the border. The average gap is 7.6 percentage points in absolute value.

Our main estimating equation is a standard difference-in-differences specification which exploits the expansion of KCC as a source of time variation and the differences in language barriers across cells as a source of cross-sectional variation as follows:

$$y_{idt} = \alpha_i + \alpha_{b(i)t} + \alpha_{dt} + \beta \left(\frac{O_i^{\text{ns}}}{O_i}\right) \times Post_t + \lambda_t \left(\frac{O_i}{N_i}\right) + \Gamma_t X_i + u_{idt} \tag{1}$$

The subscript i identifies 10×10 km cells, and t indexes years between 2006 and 2017. $Post_t$ is a dummy equal to 1 in the period after 2007. We use 2007 as our baseline *pre*-treatment year because it pre-dates the expansion of KCC documented in Figure C.1(a) and coincides with the last Agricultural Input Survey before such expansion.

The main coefficient of interest is β, which captures how language barriers between farmers and KCC advisors affect the impact of KCC on the outcome variables. We measure language barriers as the share of official language speakers (O_i) who do not speak the official language of their state (O_i^{ns}), e.g. Gujarati speakers in Hindi-speaking

[2] KCC calls are reported at the subdistrict level. We superimpose the map of subdistrict boundaries with the 10×10 km cell grid and assign calls proportionally to all cells whose centroid is contained within a subdistrict. AIS data on technology adoption and ICRISAT data on crop yields are at the district-crop level. We compute the share of land farmed with a given agricultural technology in a given cell as the sum of the district-level measures of technology adoption for each crop, weighted by the cell-level share of land farmed with each crop according to the FAO-GAEZ data in 2000 (Fischer et al., 2008). We use a similar neutral assignment rule to map crop yield information from district-level to cell-level. Appendix A explains this assignment rule in detail and validates our measure against two household surveys with information on cultivation practices. Finally, data on the Enhanced Vegetation Index (EVI), which proxies for changes in productivity, is reported at the village level and sourced from Asher and Novosad (2020). We superimpose the map of village boundaries with the 10×10 km cell grid and assign to each cell the average vegetation index across villages whose centroid is contained within a cell.

[3] Other topics that farmers consistently ask about include weather forecasts, access to credit products and government schemes, market price information. See Appendix B for more details.

Rajasthan.[4] In all specifications we control for the share of local population that speak any of the official languages of India ($\frac{O_i}{N_i}$) interacted with time fixed effects. This ensures that the relevant variation identifying β comes exclusively from the *composition* of official languages in the local population, and not from the share of individuals that do not speak any of the official languages.

All specifications include cell fixed effects (α_i), as well as common subdistrict border fixed effects interacted with year fixed effects ($\alpha_{b(i)t}$) so to compare geographically close cells on the opposite side of state borders.[5] We also control for district-specific trends (α_{dt}) and a set of baseline cell characteristics (X_i), including the share of area farmed under the 10 main crops of India, interacted with year fixed effects. Standard errors are clustered at subdistrict level (1,872 in our sample) to account for geographical correlation across cells within the same administrative unit. We weight regressions by cell population.

The main identification assumption is that, conditional on the covariates included in equation (1), the share of official language speakers that do not speak the state official language in a given cell is independent of u_{idt}. We provide an indirect test of conditional independence by looking at the correlation of language barriers with observable cell characteristics at baseline, controlling for the share of official language speakers in each cell and the same set of fixed effects as equation (1). In section III we also present event studies showing year-by-year estimates of the effect of language barriers on the outcomes of interest to test for pre-existing trends.

The results of the balance test are reported in Table II. Coefficients indicate differences in cell characteristics for one standard deviation difference in non-state language speakers. The share of non-state language speakers among official language speakers is uncorrelated with most of the observable cell characteristics, including population, agricultural employment share, literacy rate, average crop suitability, connection to the power grid, terrain ruggedness, and presence of a school or a hospital. We also test for differences in the diffusion of the main religions of India (Hindus and Muslims) and the ethnic composition of cells in our sample as captured by the share of local population that belong to "scheduled castes". Scheduled castes identify historically discriminated communities outside of the mainstream caste system. We find no significant differences in religion or caste composition. This is consistent with the fact that, differently from other settings, language differences can exist within groups with similar ethnic composition in India. Finally, the share of non-state language speakers is uncorrelated with infrastructural determinants of KCC access, such as the availability of telephone landlines and the share of area covered

[4]We define state vs non-state language speakers based on the answer to the 2011 Population Census question about an individual's mother tongue. The mother tongue is defined as "the language spoken in childhood by the person's mother to the person".

[5] We construct these fixed effects as follows. First, we identify, for each cell i, the closest point on the state border. Every point on the state border is also a border between two subdistricts, one on each side of the state border. Common subdistrict border fixed effects capture all cells whose nearest border point is shared by the same subdistrict pair.

by the 2G mobile phone network. Out of 26 covariates, only the presence of a post office at baseline shows a statistically significant correlation with language barriers. We include this variable among the controls (X_i) in all specifications.

A potential concern with our identification strategy is that official language speakers that do not speak the official language of their state are also less exposed to other government programs that were introduced contemporaneously to KCC. This correlation could exist either because such programs were only offered in the official language of their state or because their roll-out differentially targeted areas with higher language barriers. While we are not aware of contemporaneous programs offered only in state languages, in Table II we report the correlation between the share of non-state language speakers and three proxies for major government programs that were introduced in the first decade of the 2000s. These include: the rural electrification program launched in 2005 (Burlig and Preonas, 2021), the village road program launched in 2000 (Asher and Novosad, 2020), and the SMIS program financing the construction of mobile phone towers in rural areas launched in 2007 (Gupta et al., 2019). We find that cells with a higher share of non-state language speakers experienced similar increases in their probability of accessing the electrical power grid or to be connected via a paved road between the last two Census years (2001 and 2011). We also find that the share of non-state language speakers does not predict the planned construction of mobile phone towers under the SMIS program.

III Main Results

III.A Farmers' calls to KCC

We start by documenting the impact of language barriers on calls made to KCC. Column (1) of Table III reports the results of estimating equation (1) when the outcome variable is the total number of calls to KCC per 100 farmers. The magnitude of the estimated coefficient indicates that areas with a one standard deviation higher share of non-state language speakers (0.225) recorded about 0.9 fewer calls to KCC per 100 farmers per year in the period after the introduction of KCC. This corresponds to 37% less calls than the average cell in our sample.

In column (2) we focus on calls about agricultural technologies such as seed varieties, pesticides, fertilizers and irrigation systems. The results confirm the negative impact of language barriers on the amount of information received by farmers on modern agricultural technologies. We also find effects that are similar in magnitude for non-technology calls in column (3). Overall, these findings are consistent with the existence of an underserved demand for information on farming techniques by Indian farmers, and indicate that language barriers can significantly hinder their ability to access such information.

III.B Adoption of modern agricultural technologies

Next, we investigate the effect of language barriers on adoption of agricultural technologies. Our measures of technology adoption are sourced from the Agricultural Input Survey (AIS) of India, which is conducted at 5-year intervals. Our main empirical analysis focuses on the last four waves of the AIS, which occurred between 2002 and 2017. We consider the 2002 and 2007 waves as the pre-KCC period, and the 2012 and 2017 waves as the post-KCC period.

Our main measure of technology adoption is the share of land farmed with HYV seeds. These are hybrid seeds developed via cross-breeding in order to increase crop yields. They combine desirable characteristics of different breeds, including improved responsiveness to fertilizers, dwarfness, and early maturation in the growing season. HYV seeds have been available in India since the Green Revolution (the IR8 rice, flagship of the Green Revolution, was introduced in 1966), but new varieties are constantly developed and introduced in the Indian market. In the period between 2002 and 2013, 47 new varieties of different oil seeds, cereals and vegetables including rice, groundnut, wheat, millet, soy and cotton were introduced. Despite their early introduction and rapid adoption in many areas of the country, a large share of the Indian agricultural land is still not farmed using HYV seeds.

Columns (4) to (6) of Table III report the results of estimating equation (1) when the outcome variables are the share of land farmed with a given technology. We find that areas with a one standard deviation higher share of non-state language speakers experienced 0.4 percentage points lower share of area farmed with HYV seeds in the post-2007 period. This corresponds to a 1.3% decrease for the average cell in our sample. One important characteristic of HYV seeds is that they are highly respondent to fertilizers and, to attain their full potential, they require a reliable source of irrigation (Dalrymple, 1974). Thus, we expect adoption of HYV seeds to increase farmers' demand for these complementary inputs. Indeed, we find that areas with a one standard deviation higher share of non-state language speakers experience a 0.7% lower increase in fertilizers and a 1.4% lower increase in irrigation.

The magnitude of our estimates indicates that HYV adoption is 0.07 standard deviations higher in areas without language barriers than in areas with full language barriers. We can compare these magnitudes to those documented in previous studies. In particular, Cole and Fernando (2020) uses an RCT to study the effect of a mobile-based service of agricultural advice on adoption of agricultural practices by cotton farmers in India. The authors find that treated farmers increase adoption of recommended seeds by 0.09 standard deviations. This effect is comparable to what we find in our setting, although the nature of the treatment is different and so this comparison should be interpreted with caution.

III.C Agricultural productivity

Finally, we study the impact of language barriers on agricultural productivity. Our main measure of productivity is average crop yields from ICRISAT.[6] The ICRISAT dataset reports average yields for all major crops farmed in India at the district-year level. We bring the data at the cell-level exploiting the same neutral allocation rule used to map technology adoption data from AIS (See Appendix A for a detailed discussion). In Section IV we also show that our results are robust to estimating our main specification at the district-crop level instead of at the cell-level.

The main results on agricultural productivity based on the ICRISAT based measure of crop yields are reported in column (7) of Table III. The estimated effect of language barriers on yields is negative and statistically significant. Its magnitude implies that areas with a one standard deviation higher share of non-state language speakers experience a 0.6% lower increase in crop yields.

As an alternative measure of productivity, we also use the Enhanced Vegetation Index (EVI), an index of intensity of vegetation cover estimated by the US Geological Survey using the Moderate Resolution Imaging Spectro-radiometer (MODIS) aboard NASA's Earth Observing System-Terra satellite. Vegetation indices such as EVI exploit plant reflectance of electromagnetic radiations to quantify vegetation greenness in an area, whose spatial distribution is estimated from satellite images.[7]

In Table C.1 we report additional results using three satellite-based measures of yields: the difference between the maximum value of EVI observed during the agricultural season and the average value observed at the beginning of the season (EVI^{Delta}), the maximum (EVI^{Max}) and the cumulative ($EVI^{Cum.}$) values of the vegetation index observed during the relevant agricultural season of each area.[8] Coefficient estimates on language barriers obtained using satellite-based measures of yields are negative but not precisely estimated.

III.D Event-studies

In Figure II we report period-by-period estimates of the effect of language barriers on the three main outcomes of interest: calls to KCC, adoption of HYV seeds and average crop yields. The event studies for HYV adoption and crop yields show lack of pre-existing

[6]The original sources of the ICRISAT data are the Directorate of Economics and Statistics (DES) and Directorate of Agriculture of Indian States.

[7] Remote sensing has been used to estimate crop yields via satellite data since the 1970s (see Barnett and Thompson (1982) for a review of early studies). Vegetation indexes such as EVI have been shown to perform well in the estimation of crop yields: see Son et al. (2014) for an application to rice yields in Vietnam and Kouadio et al. (2014) for an application to wheat yields in Western Canada. See Asher and Novosad (2020) and Asher et al. (2021) for recent applications of the EVI as a proxy for agricultural productivity in India.

[8]By measuring changes in vegetation from the sowing period (when the land is uncultivated) to the moment of peak vegetation, EVI^{Delta} allows to partially account for differences in the underlying non-agricultural vegetation across areas, such as forest cover.

trends and yearly point estimates that become negative after the expansion of Kisan Call Centers in the period post 2007. Farmers calls to KCC are only observed after the introduction of KCC and, thus, do not allow us to test for pre-existing trends. For this outcome, we use 2009 as reference period because it is the first year with enough calls to allow us to estimate the effect of language barriers (see Figure C.1(a)). Still, we observe an increase in the effect of language barriers on calls over time in the post period.

III.E Discussion of magnitudes

The estimates in Table III indicate that areas with one standard deviation higher share of non-state language speakers experience 0.9 less calls per 100 farmers per year and a 0.4 percentage points decline in the share of land under HYV seeds. Because the measure of technology adoption is not reported at the farmer level, but as a share of cultivated area in a given location, one cannot use our estimates to compute the elasticity of number of farmers adopting HYV seeds in response to one additional call to KCC. Under the strong assumption that land is equally distributed across farmers in each cell, the 0.4 percentage points relative decline in the share of farmers using HYV seeds would correspond to 18 less farmers adopting HYV seeds. For the average cell, 0.9 less calls per 100 farmers per year corresponds to 46 less calls per year. The ratio of adoption to calls would therefore imply an elasticity of $18/46 = 0.39$.

The magnitude of these effects implies that KCC calls strongly affect farmers' decisions. To investigate farmers' perception of the information they received via KCC we rely on the results of a survey of 458 KCC users presented in Gandhi et al. (2017). Three stylized facts emerging from the survey are important to provide context for our results. First, users consider technical information on seed varieties the most important type of information for farmers among those provided by KCC. Second, about half of the callers (55%) report that technical information received from KCC was helpful in improving profits or performance. Finally, when asked whether KCC advisors understand and respond in the farmer's language, 82% agree with this statement, while the remaining 18% disagrees, suggesting that language frictions are at work in our context for a meaningful fraction of KCC users.

The impact of calls on adoption depends not only on the direct effect of KCC on callers, but also on the indirect effects of information spillovers from callers to non-callers. Previous literature has shown that spillover effects can be substantial. Cole and Fernando (2020) show that, when accounting for spillovers, the magnitude of the response of adoption of recommended seeds to the information shock increases from 0.09 to 0.11 standard deviations (22%). In our setting, we do not observe outcomes at the farmer-level, nor the information network of each farmer. Thus, any spillover effects are subsumed in the total effect of language barriers on technology adoption. Survey evidence presented in Gandhi et al. (2017) shows that farmers calling KCC are selected in terms of personal characteristics. Callers tend to be relatively more educated, the majority of them having completed secondary education. Existing evidence shows that more educated farmers are also more likely to be part of the social network of other farmers (Varshney et al., 2022), and that seeding information with a selected group of individuals that are central in the local network can be a powerful tool to disseminate information within a community (Conley and Udry, 2010; Beaman et al., 2021; Banerjee et al., 2018).

IV Robustness tests

We perform a set of robustness test of the main results presented in Table III. In the main analysis we focus on cells located within $50km$ from state borders. In Table C.2 we show that estimates are robust to using distances to state borders ranging from $40km$ to $60km$ (at intervals of $5km$). In the main analysis, standard errors are clustered at the subdistrict level. In Table C.3 we show that the main coefficients of interest remain statistically significant when standard errors are calculated accounting for spatial correlation at different distance thresholds using the correction proposed in Conley (1999).

Language barriers might mask differences in other characteristics. To further attenuate this concern, we augment our main specification with majority language fixed effects and majority religion fixed effects. Majority language is defined as the language spoken by the largest share of population in a cell. Majority religion is defined as the largest religious group in a cell. The results of this robustness test are reported in Table C.4. The estimated coefficients capturing the effects of language barriers on calls, HYV adoption and yields are stable in magnitude and remain statistically significant after including these additional controls.

Technology adoption and crop yields are measured at the district-crop level and not at the cell level. As a robustness test, we replicate our analysis at the district-crop level, focusing on districts that are geographically adjacent to state borders. Figure C.2 shows the sample used, and Table C.5 Panel A reports the results. We find negative and statistically significant effects of language barriers on HYV adoption and yields.[9] Magnitudes are larger than those observed in the cell-level specification: districts with a one standard deviation larger share of non-state language speakers experience 4% larger declines in the share of land farmed with HYV and agricultural yields. The estimated coefficients should be interpreted with caution as bringing the analysis at district level leads to larger and significant differences in baseline characteristics across districts with different initial shares of non-state language speakers. In Table C.5 Panel B we report the results obtained replicating our analysis at the subdistrict level. As shown, results are robust to this alternative unit of observation, although the effect of language barriers

[9] The effect on calls cannot be estimated with this specification because a large share of call-level data does not report information on the crop the farmer asks information about.

on calls becomes smaller in magnitude in this specification.

Because we focus our analysis on state-border cells, a potential concern is that farmers that do not speak the official language of their state might cross the border to access the service in another language. We think this channel is unlikely to be relevant. Farmers' calls to the national phone number of KCC are redirected to the state offices based on the location where the phone of the caller was registered, and not on the location of the cellphone tower that transmits the call. While we can not rule out the possibility that farmers obtain SIM cards from other places, doing so would involve producing false documentation. Moreover, the imposition of extra charges (termed as "roaming charges") when operating a cell phone from a different state than the state of residence implies that such a strategy is going to be costly.

Finally, we should discuss the issue of multilingualism, and in particular the ability of non-state language speakers to also speak the official language of the state in which they live. We use the 2011 Population Census data to calculate the diffusion of bilingualism among the population of interest for our study. The data show that only 8.26% of individuals whose first language is not the official language of the state report speaking it as a second language. While this figure is observable only at the aggregate level - so we cannot use it to explicitly study whether the effects differ according to the local diffusion of bilingualism - its low level suggests a limited relevance of the issue in our context. In addition, bilingualism might not fully capture the ability of farmers to understand technical words in another language, a common complaint of farmers using KCC (Khanal, 2015).

V Concluding Remarks

Slow adoption of new agricultural technologies in developing countries has been attributed to the lack of information available to farmers (Conley and Udry, 2010; Munshi, 2004). We document that language barriers are a significant obstacle to the diffusion of such information. These results can be informative beyond the setting of our study. For example, many countries in Sub-Saharan Africa have similar agricultural employment shares and language diversity as India. As wireless telecommunication services become increasingly available in rural areas of developing countries, so do the expectations about their ability to reduce information frictions and improve productivity (GSMA, 2020). Our results suggest that the increasing amount of information available may exacerbate differences in economic opportunities between those who are able to access this information, and those who are not.

References

Adsera, A. and M. Pytlikova (2015). The role of language in shaping international migration. *The Economic Journal 125*(586), F49–F81.

Aker, J. C., I. Ghosh, and J. Burrell (2016). The Promise (and Pitfalls) of ICT for Agriculture Initiatives. *Agricultural Economics 47*(S1), 35–48.

Alesina, A. and E. L. Glaeser (2004). *Fighting poverty in the US and Europe: A world of difference*. Oxford University Press.

Anderson, J. R. and G. Feder (2004). Agricultural Extension: Good Intentions and Hard Realities. *The World Bank Research Observer 19*(1), 41–60.

Armand, A., B. Augsburg, A. Bancalari, and K. K. Kameshwara (2022). Social proximity and misinformation: Experimental evidence from a mobile phone-based campaign in india.

Asher, S., A. Campion, D. Gollin, and P. Novosad (2021). The long-run development impacts of agricultural productivity gains: Evidence from irrigation canals in india. *WorkingPaper July*.

Asher, S. and P. Novosad (2020). Rural Roads and Local Economic Development. *American Economic Review 110*(3), 797–823.

Ban, R., S. Jha, and V. Rao (2012). Who has voice in a deliberative democracy? evidence from transcripts of village parliaments in south india. *Journal of Development Economics 99*(2), 428–438.

Banerjee, A., E. Breza, A. G. Chandrasekhar, and B. Golub (2018). When less is more: Experimental evidence on information delivery during india's demonetization. Technical report, National Bureau of Economic Research.

Barnett, T. and D. Thompson (1982). The use of large-area spectral data in wheat yield estimation. *Remote sensing of Environment 12*(6), 509–518.

Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2021). Can network theory-based targeting increase technology adoption? *American Economic Review 111*(6), 1918–43.

Bridle, L., J. Magruder, C. McIntosh, and T. Suri (2020). Experimental insights on the constraints to agricultural technology adoption.

Burlig, F. and L. Preonas (2021). Out of the Darkness and into the Light? Development Effects of Rural Electrification.

Casaburi, L., M. Kremer, S. Mullainathan, and R. Ramrattan (2019). Harnessing ICT to Increase Agricultural Production: Evidence from Kenya. *Harvard University*.

Cole, S. A. and A. Fernando (2020). Mobile-izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability. *The Economic Journal*.

Conley, T. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics 92*(1), 1 – 45.

Conley, T. G. and C. R. Udry (2010). Learning about a New Technology: Pineapple in Ghana. *American Economic Review 100*(1), 35–69.

Dalrymple, D. G. (1974). Development and Spread of High-yielding Varieties of Wheat and Rice in the Less Developed Nations. Technical report, United States Department of Agriculture, Economic Research Service.

Debaere, P., H. Lee, and J. Lee (2013). Language, ethnicity and intrafirm trade. *Journal of Development Economics 103*, 244–253.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012). The Political Economy of Linguistic Cleavages. *Journal of Development Economics 97*(2), 322–338.

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review 101*(6), 2350–90.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, 1203–1250.

Eberhard, D. M., G. F. Simons, and C. D. Fennig (Eds.) (2022). *Ethnologue, Languages of the World* (Twenty-fifth ed.). SIL International.

Evenson, R. E. and D. Gollin (2003). Assessing the Impact of the Green Revolution, 1960 to 2000. *Science 300*(5620), 758–762.

Fabregas, R., M. Kremer, and F. Schilbach (2019). Realizing the Potential of Digital Development: The Case of Agricultural Advice. *Science 366*(6471).

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

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, Iinsurgency, and Civil War. *American Political Science Review 97*(1), 75–90.

Fenske, J. and N. Kala (2021). Linguistic distance and market integration in india. *The Journal of Economic History 81*(1), 1–39.

Fischer, G., F. Nachtergaele, S. Prieler, H. Van Velthuizen, L. Verelst, and D. Wiberg (2008). Global agro-ecological zones assessment for agriculture (gaez 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy 10.*

Foster, A. D. and M. R. Rosenzweig (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy 103*(6), 1176–1209.

Gandhi, V. P., N. Johnson, et al. (2017). Decision-oriented information systems for farmers: a study of kisan call centres (kcc), kisan knowledge management system (kkms), farmers portal, and m-kisan portal. *Final Report*.

Ginsburgh, V. and S. Weber (2020). The Economics of Language. *Journal of Economic Literature 58*(2), 348–404.

GSMA (2020). Digital Agriculture Maps: 2020 State of the Sector in Low and Middle-Income Countries. *London: GSMA.*

Guillouet, L., A. Khandelwal, R. Macchiavello, and M. Teachout (2021). Language barriers in multinationals and knowledge transfers. Technical report, National Bureau of Economic Research.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics 124*(3), 1095–1131.

Gupta, A., J. Ponticelli, and A. Tesei (2019). Technology adoption and access to credit via mobile phones.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014). Learning Through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics 129*(3), 1311–1353.

Jain, T. (2017). Common tongue: The impact of language on educational outcomes. *The Journal of Economic History 77*(2), 473–510.

Khanal, V. (2015). Farmers in a fix over kisan call centre's tech language. *The Times Of India*.

Kouadio, L., N. K. Newlands, A. Davidson, Y. Zhang, and A. Chipanshi (2014). Assessing the performance of modis ndvi and evi for seasonal crop yield forecasting at the ecodistrict scale. *Remote Sensing 6*(10), 10193–10214.

Melitz, J. (2008). Language and foreign trade. *European Economic Review 52*(4), 667–699.

Munshi, K. (2004). Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics 73*(1), 185–213.

National Sample Survey (2005). Report No. 499: Situation Assessment Survey of Farmers: Access to Modern Technology For Farming. Technical report, Government of India. Ministry of Statistics and Programme Implementation.

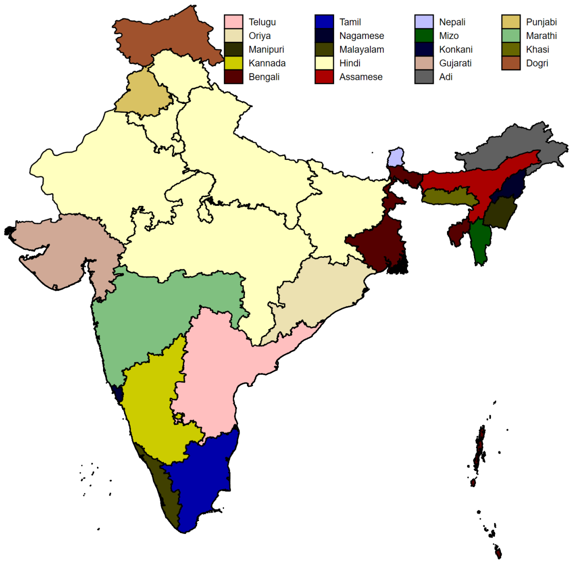
Son, N., C. Chen, C. Chen, V. Minh, and N. Trung (2014). A comparative analysis of multitemporal modis evi and ndvi data for large-scale rice yield estimation. *Agricultural and forest meteorology 197*, 52–64.

Spolaore, E. and R. Wacziarg (2009). The diffusion of development. *The Quarterly journal of economics 124*(2), 469–529.

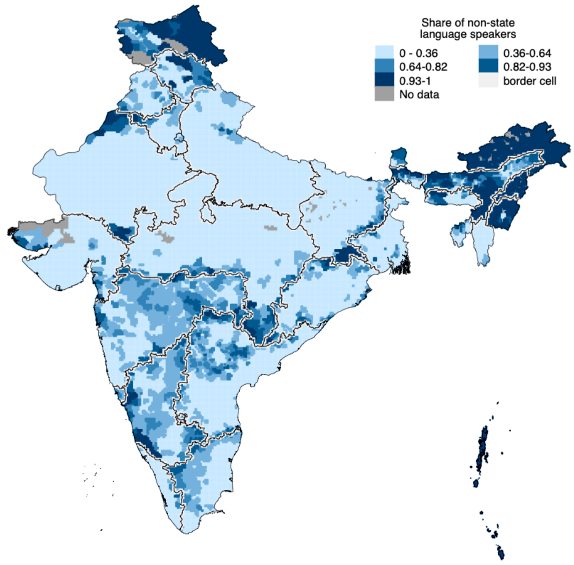
Suri, T. and C. Udry (2022). Agricultural technology in africa. *Journal of Economic Perspectives 36*(1), 33–56.

Varshney, D., A. K. Mishra, P. K. Joshi, and D. Roy (2022). Social networks, heterogeneity, and adoption of technologies: Evidence from india. *Food Policy 112*, 102360.

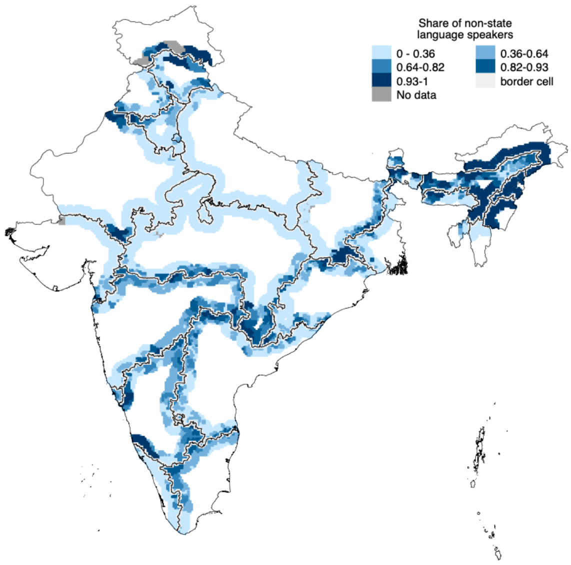
Figure I: Language heterogeneity across India



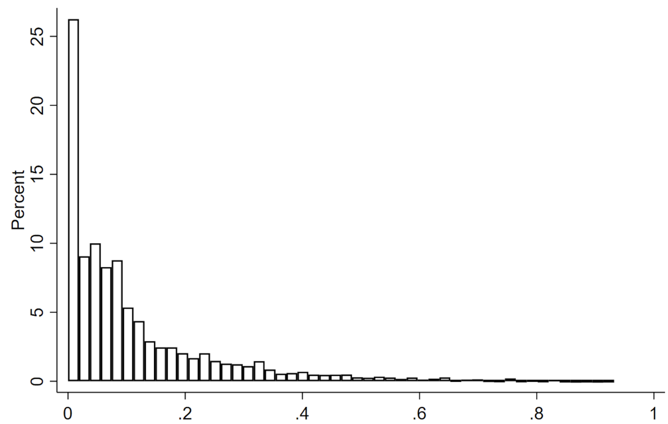
(a) Official state languages



(b) Share of non-state official language speakers



(c) Share of non-state official language speakers in cells within 50 kms of state borders

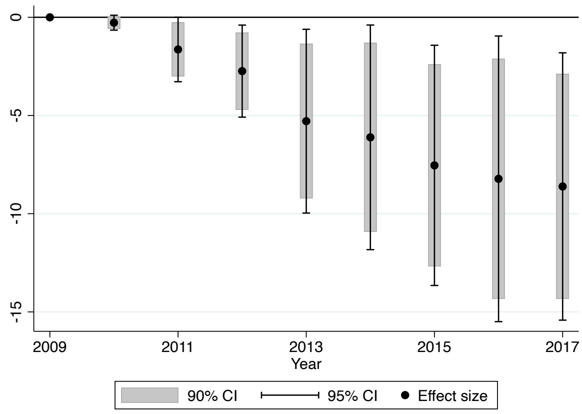


(d) Gap in non-state official language speakers across border cells

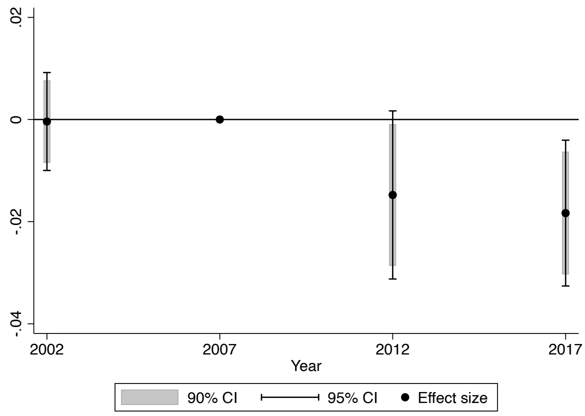
Notes: Source: 2011 Population Census of India. Speakers identified by their primary language. Panel (b) and panel (c) plot the share of non-state language speakers out of the official language speakers.

Figure II: Event Studies

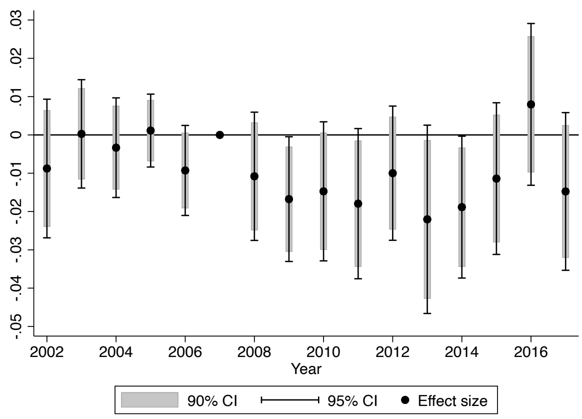
(a) Calls to KCC



(b) HYV Seeds Adoption



(c) Average Crop Yields



Note: This figure shows the event-study estimates for the three main outcomes of the paper. Panel A reports the effects on the number of calls to KCC per 100 farmers. Panel B reports the effects on the share of cell area farmed with HYV seeds. Panel C reports reports the effects on the average agricultural yield in the cell. The black bars represent 95% confidence intervals, and the gray boxes represent 90% confidence intervals.

Table I: Summary Statistics

	Mean	Standard Deviation	N	Data Source
Baseline cell characteristics				
Working Population	8,129.558	5,140.464	9,491	2001 Population Census
Share of farmers	0.627	0.243	9,491	2001 Population Census
Share of agricultural land	0.460	0.229	9,491	2001 Population Census
Share of non-state language speakers	0.130	0.225	9,491	2011 Population Census
Share of official language speakers	0.916	0.180	9,491	2011 Population Census
HYV Share (2007)	0.280	0.222	9,491	Agricultural Input Survey
Number of calls to KCC				
All calls (per 100 farmers)	2.472	29.959	104,401	Kisan Call Center
Tech calls (per 100 farmers)	1.059	14.063	104,401	Kisan Call Center
Other calls (per 100 farmers)	1.413	17.216	104,401	Kisan Call Center
Technology adoption				
HYV Share	0.299	0.222	36,151	Agricultural Input Survey
Fertilizer Share	0.324	0.208	36,151	Agricultural Input Survey
Irrigation Share	0.246	0.226	36,151	Agricultural Input Survey
Productivity				
Yield	0.405	0.222	129,843	ICRISAT

Notes: The unit of observation is a 10×10 km cell and the sample includes all border cells used for identification. The baseline cell-characteristics of working population, share of farmers and share of agricultural land are sourced from the 2001 Population Census. The share of non-state and official languages speakers are sourced from the 2011 Population Census. Baseline HYV share in 2007 is sourced from the Agricultural Input Survey.

Table II: Share of non-state language speakers and cell characteristics
Balance Test

Dependent Variable	Coefficient	Dependent Variable	Coefficient
		Availability of	
Log(Population)	-0.035	...power supply	0.006
	(0.038)		(0.007)
Ruggedness	0.005	...bus connection	-0.005
	(0.020)		(0.008)
Agri. Workers/Working Pop.	-0.000	...education facility	0.003
	(0.006)		(0.005)
Distance to nearest bank (kms)	0.058	...medical facility	0.001
	(0.074)		(0.006)
Distance to nearest town (kms)	-0.577	...post office	-0.334**
	(0.470)		(0.134)
% Area irrigated (2001)	0.005	...telephone office	0.033
	(0.003)		(0.023)
Log(crop suitability)	-0.014	...credit society	0.004
	(0.013)		(0.005)
% Land under forest	-0.005	...cooperative bank	-0.021
	(0.004)		(0.025)
Share scheduled castes population	-0.005	...communication facility	-0.002
	(0.004)		(0.008)
% Hindu population	-0.004		
	(0.011)		
% Muslim population	-0.006		
	(0.005)		
Male literacy rate (%)	-0.008		
	(0.005)		
% Area under 2G coverage (2001)	-0.006		
	(0.010)		
Δ Area under 2G coverage (2001-2011)	0.002		
	(0.012)		
Δ Access to power grid (2001-2011)	-0.014		
	(0.035)		
Δ Access to paved roads (2001-2011)	-0.009		
	(0.027)		
Planned construction of SMIS cell-phone towers	0.016		
	(0.019)		

Notes: This table reports the correlation between cell-level observable characteristics and share of non- state official language speakers in baseline year 2001. Specifically, it reports the estimated coefficient from estimating equation 1 separately for each reported dependent variable. The independent variable is normalized so that estimated coefficients can be interpreted as the difference in a given observable characteristic for a cell with one standard deviation higher share of non-state official language speakers. Standard errors clustered at subdistrict level are reported in brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .

Table III: Effects of potential access to information on calls, technology adoption and productivity

Outcome:	Calls per 100 farmers			Technology Adoption			Yield
	All	Tech	Others	HYV	Fertilizer	Irrigation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-state official language speakers (%) × Post	-4.054**	-1.947**	-2.106**	-0.016**	-0.010**	-0.015*	-0.010**
	[1.748]	[0.909]	[0.877]	[0.007]	[0.005]	[0.008]	[0.005]
Observations	104,401	104,401	104,401	36,151	36,151	36,151	129,843
R-squared	0.581	0.621	0.523	0.994	0.995	0.994	0.990
Mean	2.472	1.059	1.413	.284	.317	.242	.382
Cell f.e.	✓	✓	✓	✓	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓	✓	✓	✓	✓
District × Time f.e.	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports the estimated coefficients from equation 1. Columns 1-3 report the results for calls made to KCC; Columns 4-6 report the results on various measures of technology adoption; Column 7 reports effects on agricultural yield. Columns 1-3 include outcome data from the Ministry of Agriculture on calls made to the Kisan Call Centres and spans 2007-2017 (annual data, beginning from when data becomes available); Columns 4-6 use outcome data from the Agricultural Input Survey (AIS) from 2002-2017 (every 5 years- 2002, 2007, 2012, 2017); Column 7 uses yield data from the ICRISAT and spans 2002-2017 (annual). All column specifications include cell fixed effects, subdistrict border-time fixed effects, district-time fixed effects, and controls for the share of official language speakers, the share of area farmed under the 10 main crops in a cell, the distance to the nearest town, and post offices, all interacted with time. Standard errors clustered at the subdistrict level are reported in brackets. The control mean is the mean of the outcome variable in 2007. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Online Appendix: Not for Publication

Language Barriers, Technology Adoption and Productivity: Evidence from Agriculture in India

A Cell-level Measure of Technology Adoption and Agricultural Yields

Data on technology adoption is sourced from the Agricultural Input Survey (AIS), conducted at five-year intervals by the Ministry of Agriculture in coincidence with the Agricultural Census to collect information on input use by Indian farmers. In the survey, all operational holdings from a randomly selected 7% sample of all villages in a subdistrict are interviewed about their input use.[10] The AIS reports information on land farmed with each technology – or combination of technologies – at the district-crop level.

We construct the share of land farmed with a given agricultural technology k in a given cell i using the following neutral assignment rule:

$$\left(\frac{Area^k}{Area}\right)_{idt} = \sum_{c \in C_i} \left[\left(\frac{Area_{idc,t=2000}}{Area_{id,t=2000}} \right) \times \left(\frac{Area^k}{Area} \right)_{dct} \right] \tag{2}$$

The first element in the summation is the share of land farmed with crop c in cell i, which is observed at cell level in the FAO-GAEZ dataset and captures the initial allocation of land across crops in a given cell in the baseline year 2000.[11] The second element in the summation is the share of land farmed with technology k in district d among the land farmed with crop c. This variable captures the rate of technology adoption for a given crop in a given district and varies over time. Thus, the product of these two elements gives us an estimate of the share of land in cell i that is farmed under technology k and crop c. Summing across the set of crops farmed in cell i (C_i), we obtain an estimate of the share of land farmed with a given technology in a given cell.[12]

Similarly, we construct the measure of agricultural yield in a given cell using the following rule:

$$yield_{idt} = \sum_{c \in C_i} \left[\left(\frac{Area_{idc,t=2000}}{Area_{id,t=2000}} \right) \times \left(\frac{yield_{dct}}{yield_{dc}} \right) \right] \tag{3}$$

where $yield_{dct}$ is the production of the crop c per unit area cropped in district d in the year t. We normalize the yield across crops by dividing the district-crop-time yield $yield_{dct}$ by the average yield for that crop in the district across all years in the data $yield_{dc}$.[13]

[10] The AIS was not conducted in the states of Bihar and Maharastra before 2012. Thus, we exclude these states from our analysis.

[11] The GAEZ dataset reports information on the amount of land – expressed in hectares – farmed with a specific crop in a given cell. The data refers to the baseline year 2000. We focus on the 10 major crops by area harvested in India, namely: rice, wheat, maize, soybean, cotton, groundnut, rape, millet, sugar and sorghum. According to FAOSTAT, the area harvested with these 10 crops amounts to 135.5 million hectares and accounts for 76% of the total area harvested in India in 2000.

[12] As an example, suppose that in district d, 20% of land farmed with rice and 50% of land farmed with wheat are farmed using high-yielding variety seeds. Suppose also that 40% of land in cell i that is part of district d is farmed with rice, while the remaining 60% is farmed with wheat. Under our neutral assignment rule, we assign 38% of land in cell i to high-yielding varieties: $(0.2 \times 0.4) + (0.5 \times 0.6) = 0.38$.

[13] That is, $yield_{dc} = \frac{1}{T} \sum_{t=1}^{T} yield_{dct}$.

The within-district variation generated by our assignment rule is driven by the baseline crop composition of each cell coupled with district-crop level variation in technology adoption and agricultural yield. One potential concern with this assignment rule is that it may generate non-classical measurement error. To see this, let y_i^* be the true level of the share of adoption of a given technology in cell i, and y_i be the imputed measure. We define measurement error in estimated technology adoption by η_i such that $y_i = y_i^* + \eta_i$. After some algebra, it is easy to show that $\eta_i = \sum_{c \in C_i} s_{idc} \times (y_{idc}^* - y_{dc})$, where y_{idc}^* is the true cell-crop share under the technology, y_{dc} is the district-crop level shares obtained from the AIS data, and s_{idc} is the share of cell area farmed under crop c. Letting x_i represent our main treatment variable – i.e. share of non-official language speakers – one would then estimate $\beta = \frac{cov(y_i^* + \eta_i, x_i)}{var(x_i)} = \frac{cov(y_i^*, x_i)}{var(x_i)} + \frac{cov(\eta_i, x_i)}{var(x_i)} = \beta^* + \frac{cov(\eta_i, x_i)}{var(x_i)}$. Any correlation of η with the share of non-state language speakers will bias our estimates of how language barriers affect technology adoption.

To see how various sources of measurement error could affect our estimates, we decompose — without loss of generality — the differences in true cell-crop level shares and observed district-crop shares into a cell-specific, a district-specific, and an idiosyncratic component: $y_{idc}^* - y_{dc} = y_i + y_d + \epsilon_{idc}$. This yields the following expression for the bias in β:

$$cov(\eta_i, x_i) = cov\left(\sum_{c \in C_i} s_{idc}\ y_i, x_i\right) + cov\left(\sum_{c \in C_i} s_{idc}\ y_d, x_i\right) + cov\left(\sum_{c \in C_i} s_{idc}\ \epsilon_{idc}, x_i\right) \quad (4)$$

First, it could be that cells that differ in their share of non-state language speakers are also on a different growth trajectory. This bias is reflected in the first term on the right-hand side of the equation (4). This would happen if, for example, cells with a higher share of non-state language speakers are also cells where farmers grow crops characterized by fast technology adoption. To address this concern, in the paper we show that the share of non-state language speakers is uncorrelated with trends in technology adoption in the five years before the introduction of KCC.

Second, our estimates could be biased downwards if cells with higher non-state speakers also have smaller area farmed under the ten crops considered. This is because under the neutral assignment rule, any changes in district-level shares will be less attributable to cells with lower shares of farmed area. This bias is reflected in the second term on the right-hand side of equation (4). To control for this potential source of bias, our main specification controls for the share of cell's area under the ten crops considered.

In summary, measurement error will have to vary in a very particular way across time, technology and crops to explain our findings. Moreover, the error will have to also vary across spatially adjacent cells that share the same subdistrict borders.

A.A Validation of Technology Adoption Measures and Correlation between Yield Measures

In this section, we validate two of the measures of technology adoption (adoption of HYV seeds and irrigation) using alternative datasets that are publicly available to researchers and that contain information on technology adoption at the village level. In addition, we also investigate the correlation between our main measure of productivity and the three satellite-based proxy measures of productivity discussed in the text.

Information on the use of HYV seeds at village level is seldom available for India. One exception is the ICRISAT Village Dynamics in South Asia (VDSA) dataset, which is based on a household survey that collects information on cultivation practices. The data records the crops farmed by each household, the total area farmed under each crop and how much of the farmed area is cultivated with improved or HYV seed varieties. The finest geographical unit of observation in the VDSA data is a village. The survey covers 17 villages across the five states of Andhra Pradesh, Gujarat, Karnataka, Madhya Pradesh and Maharashtra in 2012 with non-missing information on HYV seeds.[14]

We use information in the VDSA data to calculate the total area farmed in each village under a given crop as well as how much of that area is cultivated using HYV seeds. Similarly, we use the share of area farmed with a given crop in a given cell using the data from the Agricultural Input Survey and the methodology described above. We then map each 10×10 km cell to VDSA villages based on village centroids. This provides us with 30 observations at the cell-crop level for which we observe HYV adoption in both sources. Figure A.1 shows that our measure is positively correlated with the VDSA data at village level: the slope of the line is 1.06 and statistically significant ($t = 4.33$).

We also validate our measure of irrigation using information available in the Village Census of India 2001. The Village Census reports information on area of land irrigated for all Indian villages for the year 2001. We construct a measure of share of irrigated land area for each of our 10×10 km cell by assigning villages to cells based on the geographical coordinates for the centroid of the village. We compare our measure of share of cell area irrigated in the year 2001 against the one reported in the village census data. This provides us with 25,017 observations at the cell level for which we observe share of irrigated land in both the Village Census and with our measure. Figure A.2 shows that our measure is positively correlated with the Village Census measure: the slope of the line is 1.1 and statistically significant ($t = 43.75$).

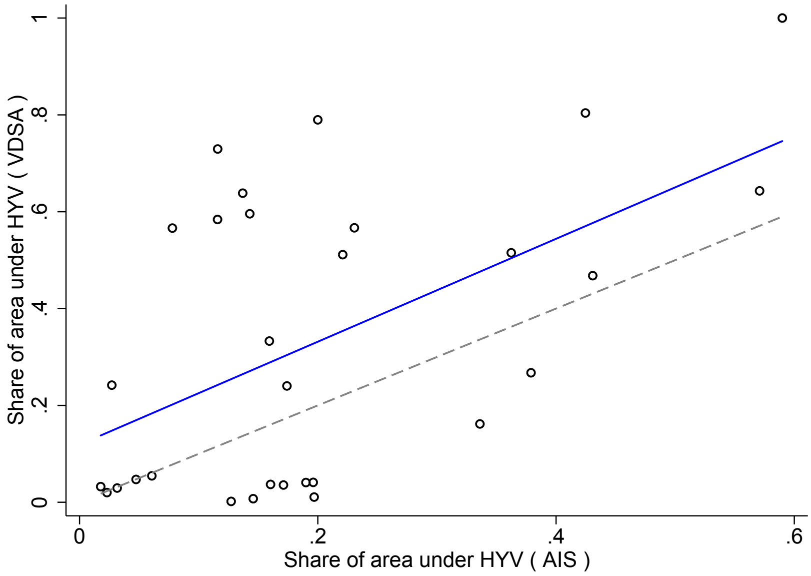
Next, we investigate the correlation between our main measure of productivity - average crop yields from ICRISAT - and the three satellite-based proxy measures of productivity based on the Enhanced Vegetation Index (EVI^{Delta}, EVI^{Max} and $EVI^{Cum.}$).

[14] A potential alternative data source on the use of HYV seeds is the Tamil Nadu Socioeconomic Mobility Survey (TNSMS) conducted by the Economic Growth Center at Yale University. One issue with the TNSMS is that it does not provide village identifiers like VDSA.

Figure A.3 considers the correlation at the district level, while Figure A.4 at the cell level. In all figures we report both the linear fit (conditional on unit of observation and time fixed effects) and the non-parametric visualization of the correlation, obtained by grouping units of observation into equal-sized bins based on their EVI value.

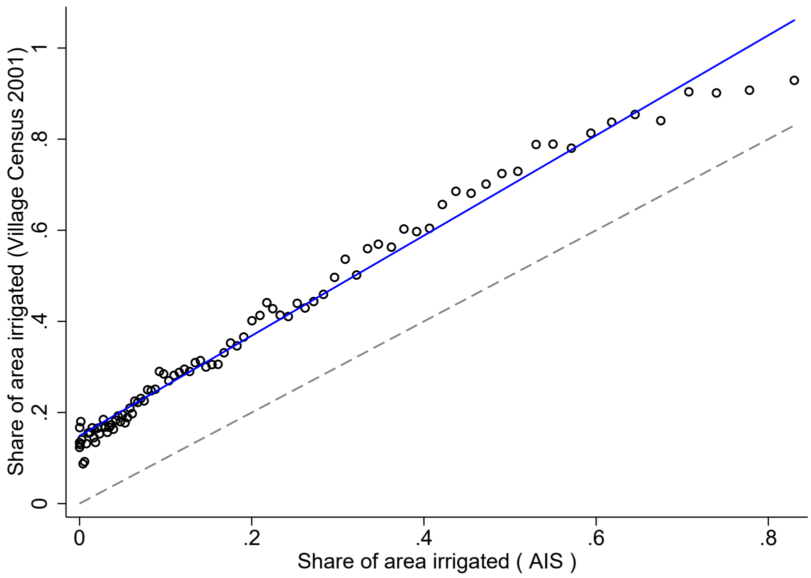
As shown, irrespective of the EVI measure and unit of analysis used, the average crop yields from ICRISAT and satellite-based measure of yields are highly positively correlated, with linear slope coefficients statistically significant at the 99% confidence level in all cases (t-statistic value ranging between 6.71 and 6.88 at the district level, and between 20.70 and 24.03 at the cell level).

Figure A.1: Data Validation: HYV Adoption



Notes: The graph reports the share of crop area under HYV as calculated from ICRISAT VDSA (Village Dynamics in South Asia) micro data against the share of crop area under HYV seeds as calculated from AIS (Agricultural Input Survey). Each dot represents a cell-crop observation for the two measures of share of area under HYV seeds in 2012. The figure has 30 observations and the slope of the line is 1.06 ($t = 4.33$). The dashed gray line is the 45 degree line.

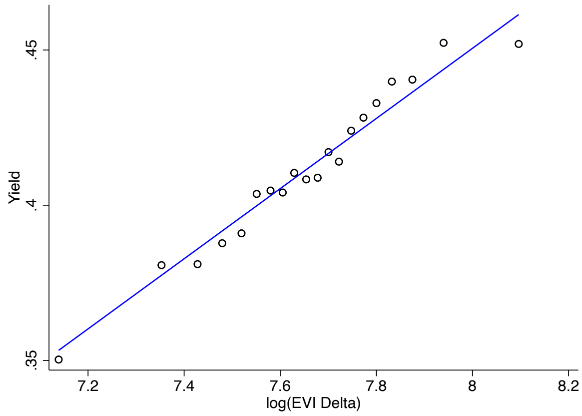
Figure A.2: Data Validation: Share of irrigated area



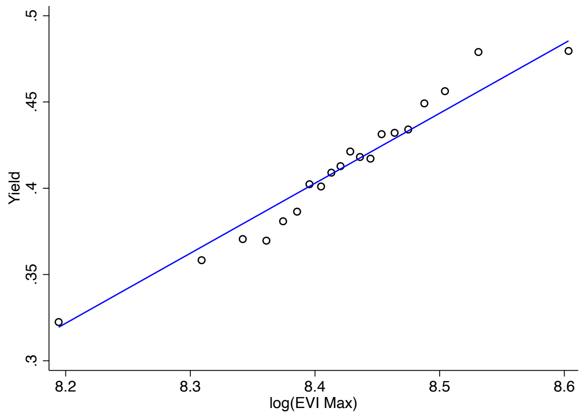
Notes: The graph reports the share of cell area under irrigation as calculated from Villages Census of India 2001 against the share of cell area under irrigation as calculated from AIS (Agricultural Input Survey) 2001. Each dot has 1% of observation based on the share of irrigated area measured through AIS and represents the average of the two measures of share of area under irrigation in 2001. The slope of the line is 1.1 ($t = 43.75$). The dashed gray line is the 45 degree line.

Figure A.3: Correlation between different productivity measurements: District level

(a) Correlation between Yield and $log(EVI^{Delta})$



(b) Correlation between Yield and $log(EVI^{Max})$



(c) Correlation between Yield and $log(EVI^{Cum.})$

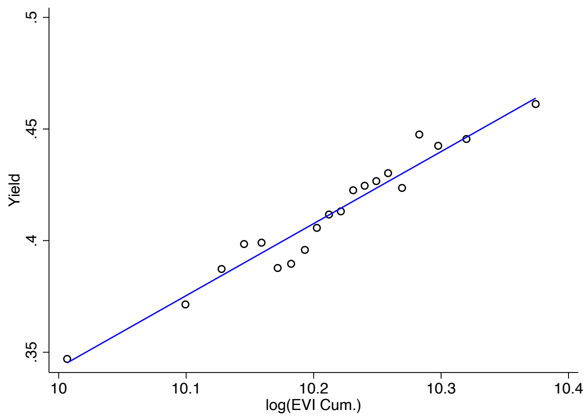
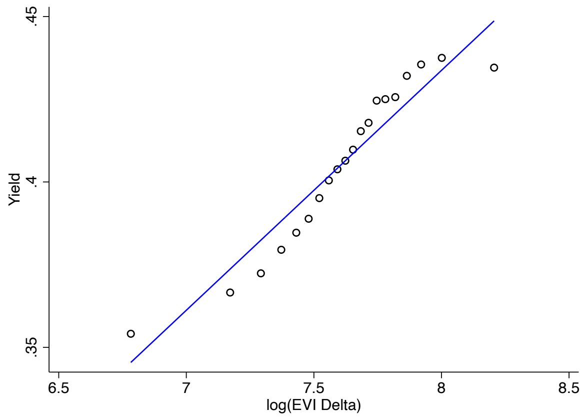
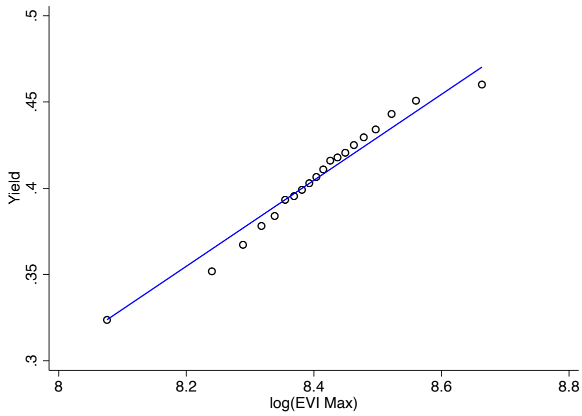


Figure A.4: Correlation between different productivity measurements: Cell level

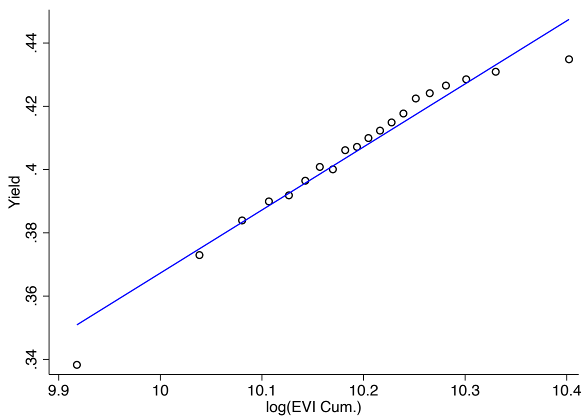
(a) Correlation between Yield and $log(EVI^{Delta})$



(b) Correlation between Yield and $log(EVI^{Max})$



(c) Correlation between Yield and $log(EVI^{Cum.})$



B Classification of KCC calls

In this section we provide examples of calls on agricultural technologies made by farmers to the Kisan Call Centres (KCC). We classify as calls about agricultural technologies those in which farmers ask questions about seeds, fertilizers, pesticides and irrigation. We extract this information from farmers' queries ("QueryText") and agronomists' answers ("Answer").

Panel A of Table B.1 refers to calls about seeds. These include (i) calls asking directly about hybrid varieties related to a crop and (ii) queries or answers about specific high-yielding seed varieties. The questions are crop-, period- and area-specific. In the examples shown, farmers call from Haryana and Punjab, two of the country's major wheat and rice producing states, respectively, to ask about high-yielding seed varieties at the beginning of their respective growing seasons, October and June.

Panel B refers to calls about fertilizers. We classify as calls on fertilizers: (i) calls seeking general information on fertilizer dosage; (ii) calls directly asking remedies for nutrient deficiencies in crops; (iii) queries or replies based on required dosage of specific fertilizers, *e.g.* N-P-K or Urea; (iv) calls seeking information on plant growth regulators, seed treatment or solution to leaf drop. In many calls farmers ask about the dosage of specific fertilizers, as reported in the examples below.

Panel C covers calls about pesticides. We classify as calls on pesticides: (i) calls seeking specific information on pesticides; (ii) agronomists suggesting the use of certain pesticides like Quinalphos and Chlorpyriphos[15]; (ii) calls asking for solutions to pest infections; (iii) calls related to plant protection; (iv) inquiries about weed control. In the examples below, farmers inquire about how to response to specific pests, from leaf-folders to termites.

Finally, Panel D refers to calls about irrigation and water management. To classify calls on irrigation, we use questions from farmers seeking information on: (i) irrigation practices; (ii) water management in the field. Most of the calls concern the suitable time for particular stages of irrigation, as shown in the examples.

[15] Quinalphos is a pesticide widely used in India for wheat, rice, coffee, sugarcane, and cotton. Chlorpyriphos is a pesticide used to kill a number of pests, including insects and worms.

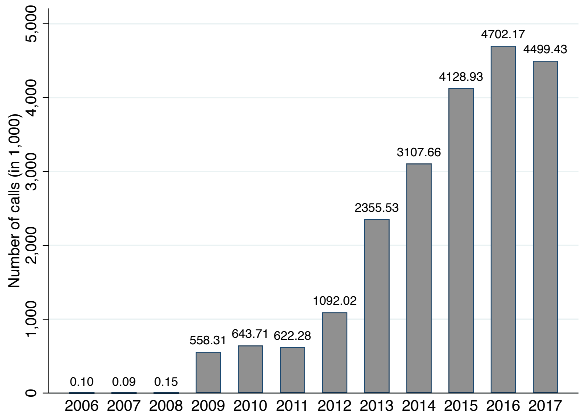
Table B.1: Examples of Calls on Agricultural Technologies

Date	State	subdistrict	QueryText	Answer
			Panel A: Calls on seeds	
2012-10-05	Haryana	Naraingarh	Information on improved varieties of wheat	w.h.-1105,w.h.d.-948, w.h-1025,w.h.-416,c.-316
2011-06-18	Punjab	Dasuya	Information on improved varieties of basmati rice	Basmati-386, Pusa Basmati No-1, Basmati-370
			Panel B: Calls on fertilizers	
2012-01-16	Chattisgarh	Manpur	To know about fertilizer in wheat at tillering	Apply 30kg.urea/acre at tillering stage
2012-02-10	Tamil Nadu	Kuttalam	Top dressing fertilizer management for rice	Apply 25 kg Urea + 15 kg Potash and 5 kg Neemcake
			Panel C: Calls on pesticides	
2012-03-14	Tamil Nadu	Tiruvallur	How to control rice leaf-folders	Spray Quinalphos at 2ml/lit
2012-02-10	Uttar Pradesh	Derapur	Termite in sugarcane	Apply Chlorpyriphos at 4lit/hac with irrigation water
			Panel D: Calls on irrigation	
2011-05-13	Haryana	Jagadhri	Time of first irrigation in cotton?	After 45 days of sowing time
2011-01-08	Rajasthan	Anupgarh	Tell me interval of time of irrigation in mustard	40-45 days

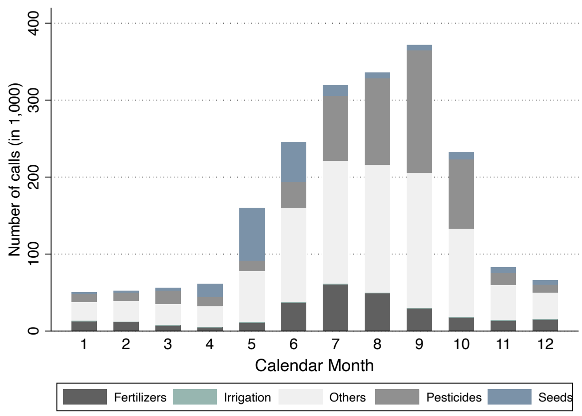
C Appendix Figures and Tables

Figure C.1: Calls to Kisan Call Centers: 2006-2017

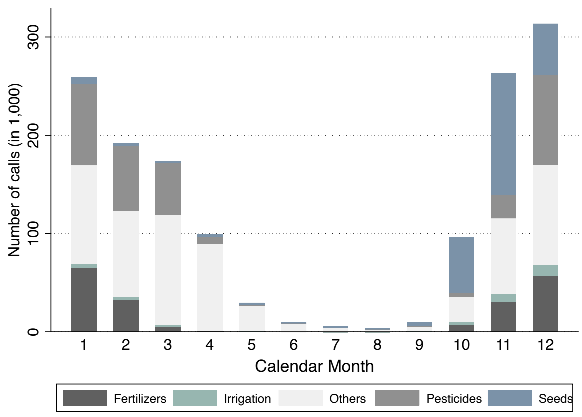
(a) Total number of calls



(b) Calls about rice (*kharif* season) by calendar month and topic

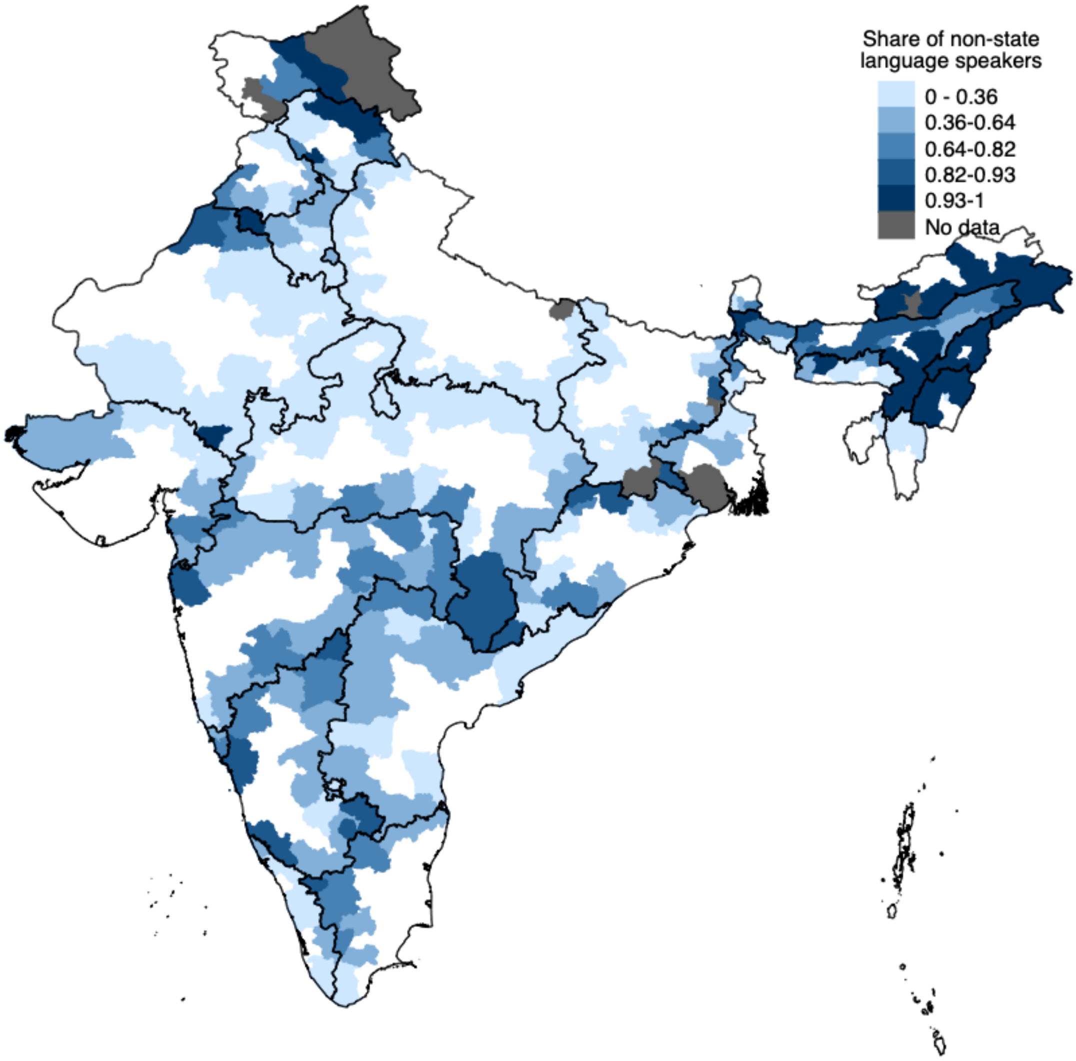


(c) Calls about wheat (*rabi* season) by calendar month and topic



Notes: Source: Kisan Call Center, Ministry of Agriculture

Figure C.2: District variation in
Non-state official language speakers



Notes: Source: 2011 Population Census of India. Speakers identified by their primary language. The figure report the share of non-state language speakers out of the official language speakers across Indian districts that are geographically contiguous to borders shared by two states.

Table C.1: Robustness: Alternate Productivity Measures

Outcome:	Yield	$\log(\text{EVI}^{\text{Delta}})$	$\log(\text{EVI}^{\text{Max}})$	$\log(\text{EVI}^{\text{Cum.}})$
	(1)	(2)	(3)	(4)
Non-state official language speakers (%) × Post	-0.010**	-0.025	-0.007	-0.003
	[0.005]	[0.023]	[0.012]	[0.014]
Observations	129,843	129,843	129,843	129,826
R-squared	0.990	0.909	0.931	0.962
Mean	0.382	7.460	8.358	10.185
Cell f.e.	✓	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓	✓
District × Time f.e.	✓	✓	✓	✓

Notes: The table reports the estimates across various measures of agricultural productivity using specification 1. Column (1) produces the estimates from Table III that using our main measure of agricultural productivity. Columns (2)-(4) report estimates using alternative measures of productivity from the Enhanced Vegetation Index (EVI), an index of intensity of vegetation cover estimated by the US Geological Survey using the Moderate Resolution Imaging Spectro-radiometer (MODIS) aboard NASA's Earth Observing System-Terra satellite. Vegetation indices such as EVI exploit plant reflectance of electromagnetic radiations to quantify vegetation greenness in an area, whose spatial distribution is estimated from satellite images. Standard errors clustered at the subdistrict level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .

Table C.2: Robustness: Alternative distances to border

Outcome:	All Calls per 100 farmers (1)	HYV Adoption (2)	Yield (3)
Panel A: Distance to border: 40 km			
Non-state official language speakers (%) × Post	-4.694**	-0.015*	-0.005
	[2.113]	[0.009]	[0.004]
Observations	88,649	30,696	112,650
R-squared	0.568	0.994	0.991
Panel B: Distance to border: 45 km			
Non-state official language speakers (%) × Post	-4.894**	-0.016**	-0.010**
	[1.907]	[0.008]	[0.005]
Observations	96,921	33,564	123,302
R-squared	0.584	0.994	0.990
Panel C: Distance to border: 50 km			
Non-state official language speakers (%) × Post	-4.054**	-0.016**	-0.010**
	[1.748]	[0.007]	[0.005]
Observations	104,401	36,151	129,843
R-squared	0.581	0.994	0.990
Panel D: Distance to border: 55 km			
Non-state official language speakers (%) × Post	-3.548**	-0.016**	-0.010**
	[1.611]	[0.007]	[0.005]
Observations	112,112	38,845	142,702
R-squared	0.580	0.994	0.990
Panel E: Distance to border: 60 km			
Non-state official language speakers (%) × Post	-3.126**	-0.016**	-0.010**
	[1.564]	[0.007]	[0.005]
Observations	118,657	41,116	150,823
R-squared	0.569	0.994	0.989
Cell f.e.	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓
District × Time f.e.	✓	✓	✓

Notes: The table reports the robustness of the main estimates from using specification 1 with alternative distances to state borders ranging from $40km$ to $60km$. Standard errors clustered at the subdistrict level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .

Table C.3: Robustness: Alternative Clustering
Conley Standard Errors

	All Calls per 100 farmers	HYV Adoption	Yield
	(1)	(2)	(3)
Non-state official language speakers (%) × Post	-4.054	-0.016	-0.010
Standard Errors (Baseline)	[1.748]**	[0.007]**	[0.005]**
Spatial Correlation, threshold: 50 km	[1.640]**	[0.005]***	[0.003]***
Spatial Correlation, threshold: 100 km	[1.788]**	[0.005]***	[0.003]***
Spatial Correlation, threshold: 150 km	[1.820]**	[0.006]***	[0.003]***
Spatial Correlation, threshold: 300 km	[1.786]**	[0.006]***	[0.003]***

Notes: The table reports the main estimates from III but adjusting for spatially clustered standard errors at 50 *km*, 100 *km*, 150 *kms* and 300 *kms*. These standard errors are reported in the table, along with the baseline standard errors clustered at the subdistrict level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .

Table C.4: Robustness to Controlling for Religion Groups and Majority Language

	All Calls per 100 farmers	HYV Adoption	Yield
	(1)	(2)	(3)
Panel A: Baseline			
Non-state official language speakers (%) × Post	-4.054**	-0.016**	-0.010**
	[1.748]	[0.007]	[0.005]
Observations	104,401	36,151	129,843
R-squared	0.581	0.994	0.990
Mean	2.472	0.217	0.354
Cell f.e.	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓
District × Time f.e.	✓	✓	✓
Religion × Time f.e.	✗	✗	✗
Language × Time f.e.	✗	✗	✗
Panel B: Controlling for Majority Religion f.e.			
Non-state official language speakers (%) × Post	-2.516*	-0.019**	-0.010**
	[1.404]	[0.009]	[0.005]
Observations	84,854	29,082	102,099
R-squared	0.592	0.993	0.987
Mean	2.209	0.217	0.354
Cell f.e.	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓
District × Time f.e.	✓	✓	✓
Religion × Time f.e.	✓	✓	✓
Language × Time f.e.	✗	✗	✗
Panel C: Controlling for Majority Language f.e.			
Non-state official language speakers (%) × Post	-3.870*	-0.021**	-0.010**
	[2.249]	[0.010]	[0.005]
Observations	104,379	36,139	129,843
R-squared	0.582	0.994	0.990
Mean	2.472	0.217	0.354
Cell f.e.	✓	✓	✓
Subdistrict border × Time f.e.	✓	✓	✓
District × Time f.e.	✓	✓	✓
Religion × Time f.e.	✗	✗	✗
Language × Time f.e.	✓	✓	✓

Notes: The table reports the robustness of the main regression specification to the inclusion of majority religion fixed effects (reported in panel B) and majority language fixed effects (panel C). Majority religion is defined as the religious group that forms the majority in a cell and is obtained from the 2011 Census data. There are 6 main religious groups across cells in our sample: Buddhists, Christians, Hindus, Muslim, Sikhs, Others. Majority language is defined as the language spoken by the majority of the population in a cell and is obtained from the 2011 Census data. There are 82 majority language groups across cells in our sample. Standard errors clustered at subdistrict level are reported in brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .

Table C.5: Robustness: Alternative Geographical Units of Observation

Outcome:	All Calls per 100 farmers	HYV Adoption	Yield
	(1)	(2)	(3)
Panel A: District-Crop as unit of analysis			
Non-state official language speakers (%) × Post		-0.133**	-0.131**
		(0.0634)	(0.0513)
Observations		4,727	24,928
R-squared		0.762	0.321
Pre-Period Mean		0.575	0.903
District × Crop f.e.		✓	✓
Border × Time f.e.		✓	✓
Crop × Time f.e.		✓	✓
Census Controls × Time f.e.		✓	✓
Panel B: Subdistrict as unit of analysis			
Non-state official language speakers (%) × Post	-0.809**	-0.013*	-0.019**
	(0.393)	(0.007)	(0.007)
Observations	17,556	4,798	17,420
R-squared	0.886	0.998	0.996
Pre-Period Mean	.714	.28	.37
Subdistrict f.e.	✓	✓	✓
Crop Share × Time f.e.	✓	✓	✓
Subdistirct border × Time f.e.	✓	✓	✓
District × Time f.e.	✓	✓	✓
Census controls × Time f.e.	✓	✓	✓

Notes: The table conducts the analysis at alternative geographical units. Panel A conducts the analysis at the district-crop level. The effect on calls cannot be estimated with this specification because a large share of call-level data does not report information on the crop the farmer asks information about. All columns include district-crop fixed effects, common state border-time fixed effects, state-time fixed effects, crop-time fixed effects, and district-level controls of share of scheduled caste population, literacy rate, total population, number of agricultural societies, and number of agricultural workers, all interacted with time fixed effects. Standard errors clustered at the district level are reported in brackets. Panel B conducts the analysis at the subdistrict level. All columns include subdistrict fixed effects, subdistrict border-time fixed effects, district-time fixed effects, share of subdistrict area farmed with 10 crops at baseline interacted with time-fixed effects, and share of subdistrict population in agriculture from the 2001 Census interacted with time-fixed effects. Standard errors clustered at subdistrict level are reported in brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1 .