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

Apoorv Gupta[†] Jacopo Ponticelli[‡] Andrea Tesei[§]

This draft: November 2022

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 call-level data from a government-sponsored call center for agricultural advice offered in different languages across Indian states. Exploiting 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. We find no impact of language barriers on agricultural productivity within five years from introduction of call centers, although small but significant differences arise in the longer run.

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. 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 diffusion of information and 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. First, India is characterized by great linguistic diversity, with 22 different languages officially recognized in its Constitution.[1] Second, 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 to assist them in their farming practices (National Sample Survey, 2005). Third, 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 territorially 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. Crucially, these contiguous areas are comparable across a large set of socio-economic and ethnic characteristics, and follow similar trends in technology adoption and productivity

[1] According to the Census of India (2011), the 22 official languages (called "Scheduled" languages) are spoken by 96.7% of the population. In addition, there are another 99 non-officially recognized languages, each spoken by at least 10,000 people. About a quarter of the Indian population is bilingual. Bilingualism is more prevalent among speakers of non-officially recognized languages and in urban areas, where English, India's main subsidiary language, is more widespread (Lok Foundation-Oxford University, 2019). Overall, the level of language fragmentation is comparable to that of Sub-Saharan Africa (Easterly and Levine, 1997).

prior to the introduction of KCC.

The empirical analysis combines 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 (see, e.g., Evenson and Gollin, 2003). Chemical fertilizers and reliable irrigation systems are key complementary inputs to maximize HYV potential. Third, we use changes in vegetation indices estimated by the US Geological Survey starting from MODIS satellite images as a proxy for changes in agricultural production. The combination of these three datasets 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 between farmers and KCC advisors experienced a 22% lower number of calls per farmer in the first decade of the program. This is consistent with language differences significantly affecting farmers' ability to access agriculture-related information. Second, we investigate the real effects of language barrier. We find that areas with higher language differences between farmers and KCC advisors experience a significantly lower increase in the adoption of certain agricultural technologies, including HYV seeds (2.3% lower adoption for a standard deviation higher language barriers), fertilizers and artificial irrigation. These effects materialized within five years from the introduction of KCC and are persistent in the long run. Third, we find no effect of language barriers on agricultural productivity in the first five years after the introduction of KCC. Statistically significant differences in agricultural productivity arise when focusing on a 10-year horizon, although these effects are small in magnitude (1% lower increase for a standard deviation higher language barriers).

Our findings speak to three important streams of the literature. First, the body of work focusing on 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, the literature has shown that 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) document how linguistic distance between two regions affected their degree of market integration during the colonial period.[2]

One challenge in this literature is that language differences tend to correlate with other factors, such as ethnic and cultural identity, preferences and group size, that may also have an impact on the outcomes of interest. Indeed, these studies often use language as a proxy for deeper cultural cleavages (Guiso et al., 2009; Ginsburgh and Weber, 2020). 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. In this sense, our findings also relate to a recent literature on the role of language differences for 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 influential micro-development 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). Compared to these studies, we 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. Previous research has highlighted the poor performance of traditional face-to-face programs, due to the lack of timely and personalized information to farmers (Anderson and Feder, 2004; Duflo et al., 2011). A key characteristic of KCC is that it gives farmers access to customized and timely information throughout the agricultural production cycle. Our results suggest that the availability of real-time agricultural advice affects farmers' adoption of modern technologies, although we detect no real effects on agricultural productivity within five years from the launch of the program, and only small effects in the long run. This is consistent with existing

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

evidence on the impact of mobile based intervention programs, which has documented – in some settings – significant effects on farmers' input choices but limited to no impact on productivity. For example, Casaburi, Kremer, Mullainathan, and Ramrattan (2019) show that sending text messages containing agricultural advice has short-term 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 households in the Indian state of 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.[3]

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 Indian 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 of such call centers, which answer calls from all Indian states. As shown in Figure I(a), KCC received less than 1,000 calls per year in the first period after its introduction. The number of calls increased substantially starting in 2009, reaching between 500,000 and 800,000 calls per year between 2009 and 2012 thanks to a large advertising campaign by the Ministry of Agriculture promoting the availability of this service among farmers. The annual number of calls increased further between 2013 and 2017, reaching about 3 million calls per year at the end of the period under study.

Our data include the universe of calls received by KCC between 2006 and 2017. The data report call-level information on the question asked by the farmer, the crop the farmer is inquiring about, a brief description of the answer provided by the KCC agronomist, and the time and location (subdistrict) from which the call was originated. Figures I(b) and (c) illustrate the type of information asked by farmers to KCC advisors reported in our data. The Figures 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. The composition of the calls is consistent with the agricultural calendar for these two crops. Rice farmers mostly ask questions about which seeds to use 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 of the calls are about how to defend the plants from

[3] For recent reviews of the broader literature on the impact of mobile phones in developing countries see Aker, Ghosh, and Burrell (2016) and Fabregas, Kremer, and Schilbach (2019).

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 the KCC agronomists. Figure II(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. However, as shown in Figure II(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 institutional feature of KCC 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 three main datasets used in the empirical analysis, which come at different levels of geographical aggregation.[4] In all our specifications, we focus on cells located within 50 *km* from state borders. The spatial distribution of our sample is displayed in Figure II(c), while the corresponding summary statistics are reported in Table I. Cells in our sample tend to be rural and specialized in agriculture. The average number of workers per cell is 8,121, with 44% of them working as farmers.

On average, around 13% of the population in our sample do not speak the official language of their state. Yet, a large number of calls to KCC originates from these cells, an average of about 15 calls per 100 farmers during the 2006 to 2017 period. Among the most frequently asked questions are those about pests and how to deal with them, what seeds to use to increase yields, fertilizers and irrigation. These categories combined, which we label as calls about technology, account for 35.6% of all calls.[5] Areas in our sample

[4] 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 (i.e., if 10 calls are originated from a given subdistrict in a year, and the subdistrict contains the centroids of 5 cells, then we assign 2 calls per cell for that year). The Agricultural Input Survey data is reported 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). 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, 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.

[5] Other topics that farmers consistently ask about include weather forecasts, access to credit products and government schemes, and market price information. Appendix B reports a description of the methodology followed to classify calls in different categories, as well as several examples.

display a significant rise in the adoption of modern agricultural technologies during the 2007-2017 decade, most notably a 4.1 percentage points increase in the share of land farmed with HYV seeds.

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 II(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. We exploit these differences to investigate the effect of language barriers on farmers calls to KCC, technology adoption and productivity.

Our main empirical specification is as follows:

$$\Delta y_i = \alpha_{b(i)} + \beta \left(\frac{O_i^{\text{ns}}}{O_i}\right) + \lambda \left(\frac{O_i}{N_i}\right) + \Gamma X_i + u_i \tag{1}$$

where i identifies a $10{\times}10$ km cell and Δy represents the change in outcomes of interest between the period before and after the introduction of KCC. We consider 2007 as our baseline *pre*-treatment year, and study changes at 5 and 10-year horizons (2012 and 2017). The choice of these particular years derives from the timing of the Agricultural Input Survey, which is carried out at 5 year intervals.

The main coefficient of interest is β, which captures the effect of language barriers between local farmers and KCC advisors. Our measure of language barriers is 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 Rajasthan. In all specifications we control for the share of local population that speak any of the official languages of India ($\frac{O_t}{N_t}$). 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 of India. We include common subdistrict border fixed effects ($\alpha_{b(i)}$), which allow us to compare geographically close cells on the opposite side of state borders.[6] We also control for a set of baseline cell characteristics (X_i), including the share of area farmed under the 10 main crops of India.[7] Finally, we cluster standard errors at the subdistrict level (of which there are 1,872 in our sample) to account for geographical correlation across cells within the same administrative unit, and we weight regressions by cell population.

Because we focus our analysis on state-border cells, farmers that do not speak the

[6] We construct these fixed effects as follows. First, we identify, for each cell i, the closest point on the state border. Notice that 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.

[7] The 10 major crops by area harvested in India are: 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.

official language of their state might cross the border in order to access the service in another language. Of course, this channel would attenuate any effect of language barriers on calls to KCC and technology adoption. Still, we think this channel is unlikely to be relevant. This is because 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.

The main identification assumption is that, conditional on the other 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_i. We provide an indirect test of conditional independence by looking at the correlation of $\left(\frac{O_i^{\text{ns}}}{O_i}\right)$ with a number of observable cell characteristics at baseline. In particular, in Panel A of Table II we run specifications analogous to equation (1), where the dependent variable is a cell characteristic sourced from the 2001 Village Survey of the Population Census of India. All coefficients indicate differences in cell characteristics for two cells that are one standard deviation apart in terms of their share of non-state language speakers. In section III we then test for differences in trends across cells with different shares of non-state official language speakers in the five years before the introduction of KCC for the main outcomes that are observable in that period.

We start by describing the results reported in Panel A of Table II. As shown, 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, presence of a school, a hospital, or a post office. We also test for differences in the ethnic composition of border cells with different share of non-state language speakers by looking at the share of local population that belong to "scheduled castes". Scheduled castes identify historically discriminated communities outside of the mainstream caste system. As shown, we find no significant differences in 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 also uncorrelated with infrastructural determinants of KCC access, such as the baseline availability of telephone landlines and the share of area covered by the 2G mobile phone network. We find a statistically significant correlation between share of non-state language speakers and average distance to the nearest town. However, the magnitude of the estimated coefficient is small, as it implies that areas with a one standard deviation higher share of non-state language speakers are, on average, about 1.3 km closer to urban centers. We include this variable among the controls (X_i) in all specifications. Notice also that we find no significant differences in the probability of having a connection via bus.

A potential concern with our identification strategy is that official language speakers that do not speak the official language of their state might also be less exposed to other contemporaneous government programs, either because such programs were only offered in the official language of their state or because their roll-out differentially targeted the areas where they live. While we are not aware of contemporaneous programs offered only in state languages, in Panel A of Table II we report the correlation between the share of non-state official 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 official 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 non-state official language speakers does not predict the planned construction of a mobile phone tower under the SMIS program.

III Main Results

III.A Farmers' calls to KCC

We are interested in studying the impact of language barriers on access to agriculture-related information, and its real effects on technology adoption and productivity. We start by documenting the impact of differences in the share of non-state language speakers on calls made to KCC, which is our measure of access to agriculture-related information. Panel A of Table III reports the results of estimating equation (1) when the outcome variable is the increase in the total number of calls to KCC per 100 farmers between 2007 and 2012 (column 1) and between 2007 and 2017 (column 2). In all columns, we focus on cells located within 50 *km* from state borders as described in section II and control for the share of official language speakers in a given cell, so that the relevant identifying variation comes from differences in the composition of state vs non-state official language speakers.

The magnitudes of the estimated coefficients indicate that areas with a one standard deviation higher share of non-state language speakers (0.225) recorded about 15 fewer calls to KCC in the first five years of the program (2007-2012) and 122 fewer calls over the first decade of the program (2007-2017). This corresponds to 31% and 22% lower calls than the average cell in our sample, in the short- and long-run respectively.[8] In the

[8] This quantification is obtained by multiplying one standard deviation in the share of non-state language speakers (0.225) by the estimated coefficients in columns (1) and (2) of Panel A (1.775 and 14.552, respectively), and then multiplying the obtained number by the average number of hundreds of farmers in each cell in our sample (37.34).

following columns, we separate calls by topic, grouping together calls regarding agricultural technologies such as new seed varieties, pesticides, fertilizers and irrigation systems. The results reported in columns (3) and (4) confirm the negative impact of language barriers between farmers and agricultural advisors 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 columns (5) and (6). Overall, the results reported in Panel A are consistent with the existence of an underserved demand for information on farming techniques by Indian farmers. The results also suggest that language barriers can significantly hinder the ability of farmers to access information on agricultural practices. Notice that, differently from the other outcomes studied in sections III.B and III.C, we cannot test for pre-trends in farmers' calls, because farmers' calls are only recorded starting with the introduction of KCCs.

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 by the Ministry of Agriculture to collect information on inputs used by farmers. Our main empirical analysis focuses on the last four waves of the AIS, which occurred between 2002 and 2017. The AIS surveys farmers about their adoption of several agricultural inputs, including the same agricultural technologies identified in the topics of calls to KCC: type of seeds – traditional vs. high-yielding varieties – chemical fertilizers, pesticides, and artificial irrigation. Our main measure of technology adoption in agriculture 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 oilseeds, 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. As reported in Table I, the average share of area farmed with HYV seeds across cells in our sample in the baseline year 2007 is 28%.

Panel B of Table III reports the results of estimating equation (1) when the outcome variable is the change in the share of land farmed with a given technology. As in Panel A, we focus on cells located within 50 km from state borders and include subdistrict border fixed effects and initial cell controls in all specifications. We start by studying the impact on changes in the share of land farmed with HYV seeds. Changes are calculated using the 2007 and 2012 waves of the AIS in column (1) and using the 2007 and 2017 waves in column (2). The estimated coefficients indicate that areas with a one standard deviation higher share of non-state language speakers experienced 0.65 percentage points lower increase in the share of area farmed with HYV seeds between 2007 and 2012, with similar effects when studying long-run changes. This corresponds to a 2.3% decrease for the average area in our sample. We test for pre-trends in HYV seeds adoption using cell-level changes between 2002 and 2007 in Panel B of Table II. As shown, the estimated coefficient is small in magnitude (-0.003) and non statistically significant, implying that cells with a higher share of non-state language speakers experienced no differential pre-trends in HYV adoption in the five-year period before the introduction of KCC.

Next, we study the effects of language barriers on changes in the share of land farmed using chemical fertilizers and artificial irrigation as additional measures of technology adoption. 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. The AIS reports the use of fertilizers and irrigation by land farmed with different types of seeds. Thus, we can estimate our main specifications splitting fertilizers and irrigation use in land farmed with HYV seeds and with traditional seeds.

In columns (3) to (6) we report results for adoption of fertilizers and irrigation in areas farmed with HYV seeds, while Appendix Table C.1 reports results for adoption of the same technologies in areas farmed with traditional seeds. As shown, the effect of language barriers on adoption of fertilizers and irrigation is concentrated in areas using HYV seeds, while we find no effects in areas farmed with traditional seeds. The magnitudes of the estimated coefficients imply a 2.2% lower increase in fertilizers and a 3.2% lower increase in irrigation in areas with one standard deviation higher share of non-state language speakers. Panel B of Table II shows no differential changes in fertilizers' adoption between 2002 and 2007, before the introduction of KCC. The coefficient on the share of non-state language speakers is small in magnitude (0.008) and non statistically significant. As for irrigation, we find that cells with a standard deviation higher share of non-state language speakers experienced a 1.3% faster adoption of artificial irrigation in the five years before the introduction of KCC. That is, pre-trends in irrigation have the opposite sign of the effects documented starting after 2007, with a magnitude of about one third of the effect documented in Panel B of Table III.

III.C Agricultural productivity

Finally, we study the impact of language barriers on local agricultural productivity. Because India lacks data on agricultural yields at fine-geographical level, we rely on satellite-derived estimates of changes in vegetation in a given location as a proxy for changes in agricultural production. In particular, we use changes in 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.[9]

We estimate equation (1) using as outcome variables three proxies of agricultural production (Asher et al., 2021). First, we use 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}). By measuring changes in vegetation from the sowing period (when the land is uncultivated) to the moment of peak vegetation, this measure allows to partially account for differences in the underlying non-agricultural vegetation across areas, such as forest cover. We also present results using the maximum (EVI^{Max}) and cumulative ($EVI^{Cum.}$) values of the vegetation index, observed during the relevant agricultural season of each area.

The results are reported in Panel C of Table III. As shown, we find small and not statistically significant effects of language barriers on the change in agricultural productivity in the early years of KCC, between 2007 and 2012. This is despite areas with lower language barriers experienced faster adoption of modern agricultural technologies in the same years, as documented in Panel B. These results could, of course, be partly explained by the fact that change in vegetation indexes are a noisy measure of agricultural productivity. However, these results are in line with previous literature documenting positive effects of extension programs on adoption of agricultural technologies but mixed results on their impact on the productivity achieved by treated farmers (e.g., Anderson and Feder, 2004).

When focusing on long-run changes, we find evidence that language barriers may influence agricultural productivity, according to the three satellite-based proxies of agricultural productivity. However, economic magnitudes are small. Focusing on our preferred measure, the estimated coefficient in column (2) indicates that areas with a one standard deviation larger share of non-state language speakers experienced a 1% smaller increases in agricultural productivity between 2007 and 2017. Again, we find no significant pre-trends for this measure in the 2002 to 2007 period, as shown in Panel B of Table II.

Overall, the results presented in Table III indicate that language barriers significantly limit access to information, with real effects on adoption of agricultural technologies. The

[9] 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.

results also suggest that language barriers may have a negative impact on agricultural productivity, though these effects are not precisely estimated and only detected in the long run.

IV Concluding Remarks

Despite considerable progress over the last half century in the development and improvement of new agricultural technologies, most agriculture in rural areas of developing countries continues to be conducted using traditional methods (World Bank, 2019). Slow adoption in these areas has often been attributed to the lack of information available to farmers about modern technologies and the best practices associated with their use (e.g., Conley and Udry, 2010; Munshi, 2004). In this paper, we emphasize one key obstacle to the diffusion of such information, represented by language barriers. We focus on India, due to its large employment share in agriculture and significant language diversity. We exploit language mismatches between farmers and agricultural advisors generated by a new phone-based government extension program to estimate the impact of language barriers on technology adoption and productivity. Our results indicate that language barriers play an important role in slowing down the modernization of agriculture in rural India. We believe that these results generalize beyond the period and sector considered. 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.

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*.

Census of India (2011). Language atlas of india.

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

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.

Economist Intelligence Unit (2012). Competing across borders: How cultural and communication barriers affect business. *The Economist*.

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.

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.

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.

Lok Foundation-Oxford University (2019). *Percentage of Indian English Speakers*.

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.

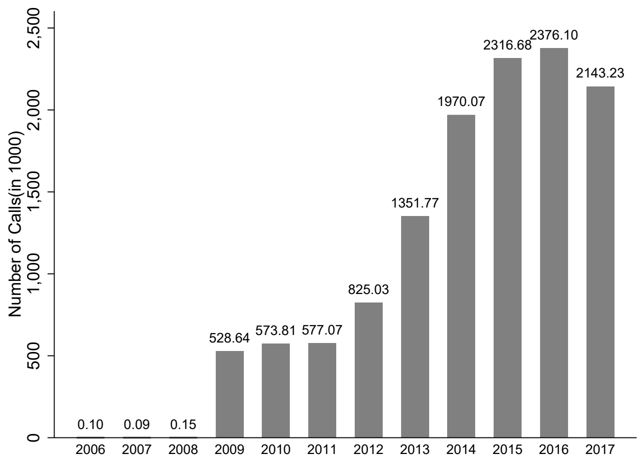
Spolaore, E. and R. Wacziarg (2009). The diffusion of development. *The Quarterly journal of economics 124*(2), 469–529.

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

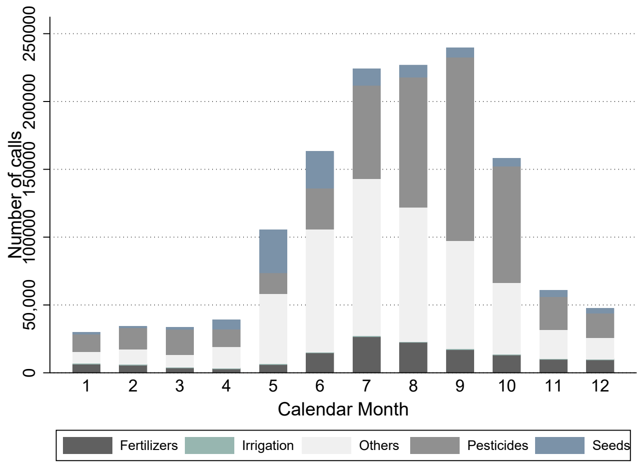
World Bank (2019). *Harvesting Prosperity: Technology and Productivity Growth in Agriculture*. The World Bank.

Figure I: Calls to Kisan Call Centers: 2006-2017

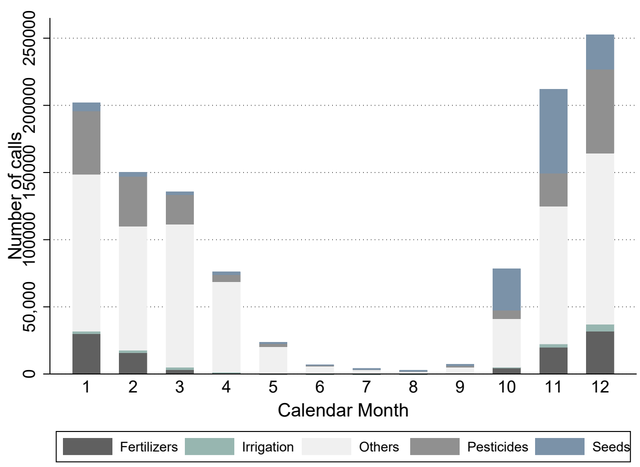
(a) Total number of calls



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

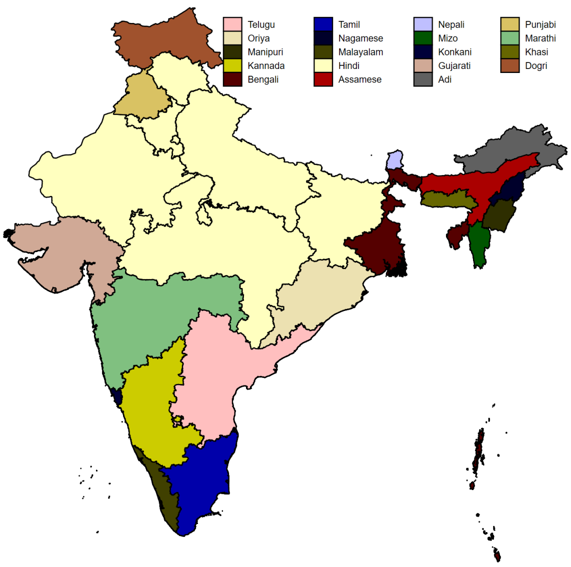


(b) Calls about wheat (*rabi* season) by calendar month and topic

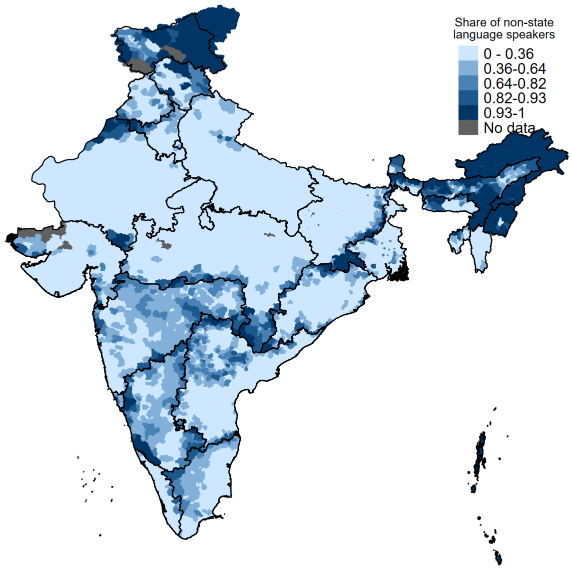


Notes: Source: Kisan Call Center, Ministry of Agriculture

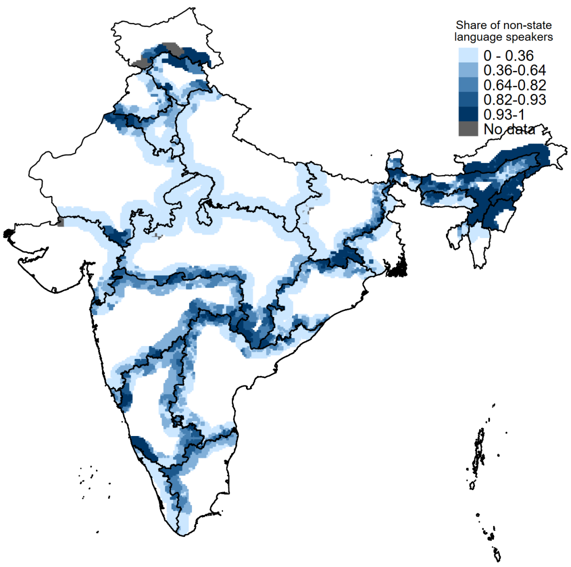
Figure II: Language heterogeneity across India



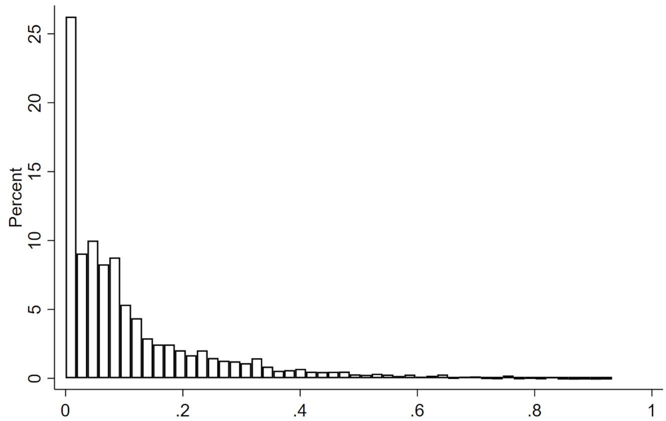
(a) Official state languages



(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.

Table I: Summary Statistics

			Mean	Std. Dev.	N
Baseline cell characteristics:					
Working Population			8121.38	5133.37	9588
Share of farmers			0.441	0.232	9588
Share of agricultural land			0.460	0.229	9588
Share of non-state language speakers			0.131	0.225	9588
Share of official language speakers			0.916	0.181	9588
HYV share			0.279	0.222	9588

	2007-2012		2007-2017		
	Mean	Std. Dev.	Mean	Std. Dev.	N
Number of calls to KCC:					
Total calls (per 100 farmers)	1.265	16.74	14.66	127.943	9588
Technology calls (per 100 farmers)	0.138	2.781	5.216	50.425	9588
Calls about other topics (per 100 farmers)	1.126	14.623	9.444	82.876	9588
Changes in technology:					
Δ HYV share	0.025	0.076	0.041	0.096	9588
Δ Fertilizer and HYV share	0.014	0.07	0.019	0.104	9499
Δ Irrigation and HYV share	0.019	0.057	0.021	0.079	9588
Changes in productivity:					
Δ log ($\text{EVI}^{\text{Delta}}$)	0.104	0.235	0.143	0.259	9021
Δ log (EVI^{Max})	0.042	0.097	0.059	0.111	9021
Δ log ($\text{EVI}^{\text{Cum.}}$)	0.016	0.078	0.025	0.095	9018

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

Panel A. Cell-level characteristics			
Dependent variable	**Coefficient**	**Dependent variable**	**Coefficient**
		Availability of ...	
Log(Population)	0.026	...power supply	0.002
	[0.026]		[0.007]
Ruggedness	0.018	...bus connection	-0.001
	[0.026]		[0.008]
Agri. Workers/Working Pop.	-0.003	...education facility	0.003
	[0.006]		[0.004]
Distance to nearest bank (kms)	0.043	...medical facility	0.004
	[0.062]		[0.009]
Distance to nearest town (kms)	-1.284***	...post office	-0.141
	[0.425]		[0.104]
% Area irrigated	0.008	...telephone office	0.010
	[0.007]		[0.027]
Log (crop suitability)	0.003	...credit society	0.002
	[0.020]		[0.005]
% Land under forest	0.005	...cooperative bank	-0.034
	[0.005]		[0.022]
Share scheduled castes population	-0.004	...communication facility	0.002
	[0.003]		[0.007]
Male literacy rate (%)	-0.007		
	[0.005]		
% Area under 2G coverage (2007)	0.015		
	[0.013]		
Δ Access to power grid (2001-2011)	0.030		
	[0.026]		
Δ Access to paved roads (2001-2011)	-0.024		
	[0.023]		
Planned construction of SMIS cell-phone towers	-0.016		
	[0.014]		

Panel B. Pre-trends in main outcomes (2002-2007)	
Dependent variable	**Coefficient**
Δ HYV Share (2002-2007)	-0.003
	[0.007]
Δ Fertilizer and HYV share (2002-2007)	0.008
	[0.008]
Δ Irrigation and HYV share (2002-2007)	0.013*
	[0.007]
Δ log [EVI$^{\text{Delta}}$] (2002-2007)	0.026
	[0.042]

Notes: Panel A reports the correlation between cell-level observable characteristics and share of non-state official language speakers in baseline year 2000 (unless otherwise indicated). 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. Panel B reports pre-trends for main outcomes during the period 2002-2007. 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

Panel A. Calls to KCC

Outcome:	Calls per 100 farmers					
	All		Tech		Others	
	07-12	07-17	07-12	07-17	07-12	07-17
	(1)	(2)	(3)	(4)	(5)	(6)
Non-state official language speakers (%)	-1.652**	-14.133**	-0.113*	-5.605***	-1.539**	-8.528*
	[0.741]	[6.047]	[0.060]	[1.712]	[0.706]	[4.723]
Observations	9,588	9,588	9,588	9,588	9,588	9,588
R-squared	0.092	0.102	0.097	0.104	0.092	0.100
Subdistrict border f.e.	✓	✓	✓	✓	✓	✓

Panel B. Technology adoption

Outcome:	$\Delta\left(Area_i^{\text{Tech.}}/Area_i^{\text{Total}}\right)$					
Technology:	HYV Seeds		Fertilizers and HYV		Irrigation and HYV	
	07-12	07-17	07-12	07-17	07-12	07-17
	(1)	(2)	(3)	(4)	(5)	(6)
Non-state official language speakers (%)	-0.029***	-0.024*	-0.022**	-0.026*	-0.039***	-0.029**
	[0.009]	[0.013]	[0.008]	[0.014]	[0.008]	[0.012]
Observations	9,588	9,588	9,499	9,499	9,588	9,588
R-squared	0.555	0.631	0.639	0.599	0.447	0.554
Subdistrict border f.e.	✓	✓	✓	✓	✓	✓

Panel C. Agricultural productivity

Outcome:	Δ log(EVI$^{\text{Delta}}$)		Δ log(EVI$^{\text{Max}}$)		Δ log(EVI$^{\text{Cum.}}$)	
	07-12	07-17	07-12	07-17	07-12	07-17
	(1)	(2)	(3)	(4)	(5)	(6)
Non-state official language speakers (%)	0.000	-0.055*	-0.006	-0.033**	-0.011	-0.026
	[0.032]	[0.033]	[0.012]	[0.014]	[0.013]	[0.018]
Observations	9,021	9,021	9,021	9,021	9,019	9,018
R-squared	0.512	0.519	0.604	0.512	0.582	0.469
Subdistrict border f.e.	✓	✓	✓	✓	✓	✓

Notes: The table reports the estimated coefficients from equation (1). Panel A reports the results on changes in calls made to KCC; Panel B reports the results on various measures of technology adoption; Panel C reports the results on various measures of agricultural productivity. Odd columns report the short-run (2007 to 2012) estimates. Even columns reports the long-run (2007 to 2017) estimates. All columns specifications include subdistrict border fixed effects, and controls for share of official language speakers, the share of area farmed under the 10 main crops in the cell, and the distance to nearest town. Standard errors clustered at subdistrict level are reported in brackets. 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

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]

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. One potential concern with this assignment rule is that it may generate non-classical measurement error. To see this, let Δy_i^* be the true growth in the adoption of a given technology in cell i, and Δy_i be the imputed measure. We define measurement error in estimated technology adoption by η_i such that $\Delta y_i = \Delta y_i^* + \eta_i$. After some algebra, it is easy to show that $\eta_i = \sum_{c \in C_i} s_{idc} \times (\Delta y_{idc}^* - \Delta y_{dc})$, where Δy_{idc}^* is the true cell-crop level growth rate under the technology, Δy_{dc} is the district-crop level growth rate obtained from the AIS data, and s_{idc} is the share of cell area farmed under crop c. Letting x_i represent

[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: $(0.2 \times 0.4) + (0.5 \times 0.6) = 0.38$.

our main treatment variable – i.e. share of non-official language speakers – one would then estimate $\beta = \frac{cov(\Delta y_i^* + \eta_i, x_i)}{var(x_i)} = \frac{cov(\Delta y_i^*, x_i)}{var(x_i)} + \frac{cov(\eta_i, x_i)}{var(x_i)} = \beta^* + \frac{cov(\eta_i, x_i)}{var(x_i)}$. 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 growth rate and observed district-crop growth rate into a cell-specific, a district-specific, and an idiosyncratic component: $\Delta y_{idc}^* - \Delta y_{dc} = \Delta y_i + \Delta y_d + \epsilon_{idc}$. This yields the following expression for the bias in β:

$$cov(\eta_i, x_i) = cov\left(\sum_{c \in C_i} s_{idc}\ \Delta y_i, x_i\right) + cov\left(\sum_{c \in C_i} s_{idc}\ \Delta y_d, x_i\right) + cov\left(\sum_{c \in C_i} s_{idc}\ \epsilon_{idc}, x_i\right) \quad (3)$$

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 3. 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 under farmed under the ten crops considered. This is because under the neutral assignment rule, any changes in district-level growth rate will be less attributable to cells with lower share of farmed area. This bias is reflected in the second term on the right-hand side of the equation 3. However, our main specification controls for the share of cell's area under the 10 crops, and therefore, the above source will not generate bias in our estimates.

Third, if any cell-crop level unobservables are correlated with our treatment variable then it could generate measurement error and potentially bias our estimates (the final term in equation 3). Notice that such an error would generate bias independent of the technology under consideration. However, the results that language barriers matter for fertilizers and irrigation only under HYV seeds (Panel B, Table III) and not under traditional seeds (Table C.1) are inconsistent with this hypothesis. 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

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.

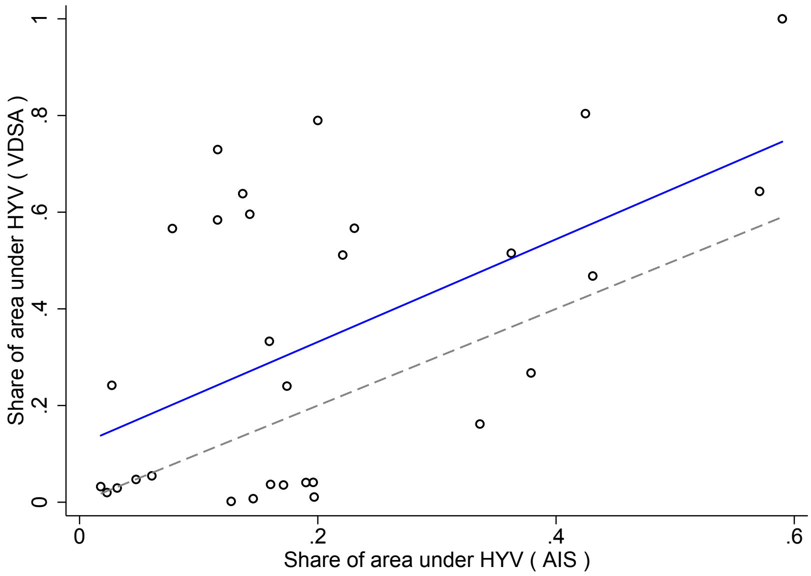
Information on the use of HYV seeds at village level is seldom available for India. One exception is the ICRISAT Village Dynamics in South Asia (VDSA) dataset, which is based on a household survey that collects information on cultivation practices. The data records the crops farmed by each household, 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.[13]

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 ($t = 43.75$).

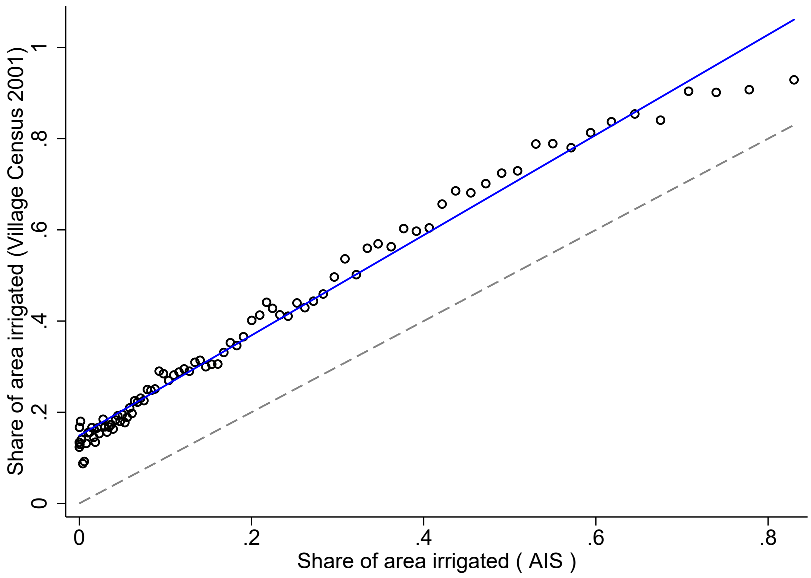
[13] A potential alternative data source on the use of HYV seeds is the Tamil Nadu Socioeconomic Mobility Survey (TNSMS) conducted by the Economic Growth Center at Yale University. One issue with the TNSMS is that it does not provide village identifiers like VDSA.

Figure A.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.

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[14]; (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.

[14] 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 Additional Robustness Table

Table C.1: Robustness: Adoption of fertilizer and irrigation in areas farmed with traditional seeds

Outcome:	$\Delta\left(Area_i^{\text{Tech.}}/Area_i^{\text{Total}}\right)$			
Technology:	Fertilizers and non-HYV		Irrigation and non-HYV	
	07-12	07-17	07-12	07-17
	(1)	(2)	(3)	(4)
Non-state official language speakers (%)	0.004	0.013	0.005*	-0.010
	[0.010]	[0.012]	[0.003]	[0.010]
Observations	9,499	9,499	9,588	9,588
R-squared	0.641	0.504	0.531	0.430
Subdistrict border f.e.	✓	✓	✓	✓

Notes: The table reports the estimated coefficient of non-state language speakers on adoption of fertilizers and irrigation in areas farmed with non-HYV seeds, using equation (1). Odd columns report the short-run (2007 to 2012) estimates. Even columns reports the long-run (2007 to 2017) estimates. All columns specifications include subdistrict border fixed effects, and controls for share of official language speakers, the share of area farmed under the 10 main crops in the cell, and the distance to nearest town. Standard errors clustered at subdistrict level are reported in brackets. Significance levels: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.