Bridging the Information Gap: Sowing the Seeds of Productivity with High-Speed 4G Internet *

Sumit Agarwal† Mohit Desai‡ Pulak Ghosh§ Nishant Vats¶

Abstract

Can high-speed internet boost information access and enhance productivity? Combining granular geographic data on the introduction of 4G with remote-sensing data on agricultural productivity, we show that the improvement in information dissemination due to the introduction of 4G leads to an increase in productivity, fertilizer consumption, and credit uptake. Our identification strategy exploiting the staggered state-level introduction of Rights of Way (RoW) policies meant to promote the growth of telecom infrastructure echoes similar results. Overall, we find that six years after the introduction of 4G internet, the annual income of agricultural households grew by 14.5%. Using detailed farmer-level internet-browsing data, we show that the introduction of 4G is related to internet adoption and acquiring agri-related information. Exploiting spatial heterogeneity in the value, reliability and accuracy of information we argue that 4G improves productivity by improving access to information. We document that the decentralized nature of internet-based information access dominates traditional call or text-based information access by circumventing frictions associated with trust in the state. While our results indicate that high-speed internet is an important tool for information dissemination, merely introducing internet infrastructure may not be sufficient. Information that is disseminated must be reliable and valuable, making internet access a complement to information generation.

*We thank John Barrios, Todd Gormley, and Margarita Tsoutsura for helpful discussions.
†National University of Singapore, Email: ushakri@yahoo.com
‡Kenan-Flagler Business School, UNC Chapel Hill, Email: mohit_desai@kenan–flagler.unc.edu
§Indian Institute of Management Bangalore, Email: pulak.ghosh@iimb.ac.in
¶Olin Business School, Washington University at St Louis, Email: vats@wustl.edu

1 Introduction

Access to information stands as a cornerstone for fostering economic development (Stiglitz (2000)). For instance, frictions that impede access to information can force economic agents to base decisions on imperfect knowledge leading to sub-optimal outcomes. Technology, particularly the widespread availability of high-speed internet, has revolutionized the way we access and consume information. The seamless flow of data facilitated by high-speed internet can alleviate frictions to information access and bridge gaps in knowledge dissemination. The existing research has focused on the impact of digital infrastructure on labor market outcomes, political economy outcomes, financial inclusion, and credit market outcomes.[1] However, we still have scarce direct empirical evidence on – (1) whether high-speed internet can affect productivity, (2) the role of high-speed internet in mitigating information frictions as a mechanism influencing productivity, and (3) the conditions under which high-speed internet effectively alleviates information frictions.

In this paper, we aim to fill this gap by studying the impact of the introduction of 4G based high-speed internet on agricultural productivity and the underlying mechanism. We investigate this question in the context of agriculture, which is well suited for a number of reasons. First, lack of information access can be detrimental in economic settings plagued by uncertainty. Agriculture in developing countries is often marked by weather-induced uncertainty (Deaton (1990), Townsend (1994)) and lack of stable insurance markets (Mobarak and Rosenzweig (2013), Cole and Xiong (2017)) making access to information an important determinant of the output and input choice (Rosenzweig and Udry (2013)). Second, existing research indicates minimal impact on agricultural productivity through traditional approaches (Fabregas, Kremer and Schilbach (2019)).[2] These traditional approaches suffer from several frictions such as lack of trust (Cole and Sharma (2017)), linguistic barriers (Gupta, Ponticelli and Tesei (2023*b*)), and timeliness (Anderson and Feder (2004)). High-speed internet through 4G has the potential to overcome these barriers by enhancing real-time information accessibility for farmers that is user-friendly and engaging (Fabregas, Kremer and Schilbach (2019)). Third, numerous governments are striving to enhance farmers' information access through

[1] See, labor market outcomes (Akerman, Gaarder and Mogstad (2015), Hjort and Poulsen (2019), Zuo (2021)), political mobilization (Manacorda and Tesei (2020)), corruption (Andersen et al. (2011), Gonzalez (2021)), confidence in governments (Guriev, Melnikov and Zhuravskaya (2021)), financial inclusion (Aker and Mbiti (2010), Suri and Jack (2016), Suri (2017), Batista and Vicente (2023)), and credit markets (DAndrea and Limodio (2023), Gupta, Ponticelli and Tesei (2023*a*)).

[2] In the absence of the internet, information is disbursed via traditional means including television, radio, phone calls, texts, extension workers or other informal means including word of mouth within communities. The relative advantage of 4G is that it improves access to high-speed internet, whereas 2G and 3G primarily improve access to texts, calls and slow internet.

the internet (Mehrabi et al. (2021)). Therefore, from a policy perspective, it is crucial to comprehend the specific conditions under which such investments are likely to yield significant returns.

We attempt to answer this question in the context of a developing economy where information frictions are acute and a majority of households rely on agriculture for their livelihood. Particularly, we focus on India where more than 40% of the population is engaged in agriculture and there is an unmet need for information. As per the 2005 National Sample Survey, 60% of Indian farmers indicated a lack of access to any information source to support their farming practices. Moreover, the Indian context provides an ideal natural experiment that allows us to disentangle the effect of the 4G based high-speed internet on agricultural productivity from other confounding variables.

We begin by exploiting the staggered introduction of 4G towers in a difference-in-differences (DiD) framework to estimate the effect of 4G on agricultural productivity. However, the exercise poses a key challenge as geographic coverage of a 4G tower is limited to a relatively small area around the tower whereas measures of agricultural productivity are typically available at a more aggregated level such as districts. Therefore, using administrative units as a unit of analysis violates the assumption of homogeneous treatment within treated units. We address this concern by employing a two-step approach. First, we divide the map of India into hexagons, each approximating the coverage area of a 4G tower. Subsequently, we use remote-sensing satellite data to create hexagon-level measures of agricultural productivity, specifically aggregating 8-day Enhanced Vegetation Index (EVI) composites to this level. Additionally, we integrate various economic, demographic, and agricultural variables into this dataset. Therefore, a key contribution of this paper is creating a granular database that can be employed in future research; his database links telecom coverage with several economic, agricultural, and demographic variables.

Our heterogeneity-robust Callaway and Sant'Anna (2021) DiD estimates exploiting the staggered introduction of 4G towers imply that six years after 4G introduction, EVI implied yield increases by 0.035 units. This increase is equivalent to 12.5% of the mean value of EVI implied yield. Economically the effect on EVI implied yield is equivalent to – (1) 1.06% increase in agricultural yields of cereal crops, (2) a 4.28% increase in the production value of cereal crops, and (3) a 3.21% increase in the production value of all crops. We estimate that six years after the introduction of 4G internet, the annual income of agricultural households grew by 14.5%. To put these numbers in context, as per the Situation Assessment Survey of India the income of agricultural households grew by 59.01% over the period from 2014 until 2021.

Our back of the envelope calculations suggest that 24.56% of the income growth of agricultural households over this period can be explained by the introduction of high-speed 4G internet.

Our empirical strategy exploits the variation across treated and not-yet-treated hexagonal units. Specifically, we compare hexagonal units that have already experienced the introduction of 4G towers with units that are slated for introduction at a later stage. A concern is that early treated units are different from units that are treated later. However, as long as these differences are time-invariant, this selection is fully accounted for by unit fixed effects. Thus, our specification does not require that the introduction of 4G BTS was randomly allocated, nor does it require that units must have the same pre-treatment characteristics. Rather our estimate would be unbiased if the treated units would have evolved the same way as the not-yet-treated units in the absence of the treatment. We substantiate this assumption by presenting the pre-trend analysis in all our empirical assessments.

There may still be concerns associated with the endogeneity in the spatial introduction of 4G towers. Prior literature has employed lightning as an instrument for the introduction of 2G/3G towers. However, lightning plays an important role in the nitrogen cycle and can contribute to the availability of nitrogen for plant growth. Therefore, such an instrument is likely to violate the exclusion restriction as the intensity of lightning can directly affect agricultural output and other economic variables in an agrarian economy.

We circumvent this issue by exploiting a plausibly exogenous variation in the roll-out of 4G technology due to the staggered state-level introduction of Rights of Way (RoW) policies. RoW policies were adopted at the state level with the intention to promote the growth of telecom infrastructure by streamlining rules and reducing the regulatory burden on telecom infrastructure companies. We first establish that the enactment of the RoW rules had a positive effect on the construction of 4G towers in those states. Specifically, the number of towers within a district increased by 600 within six quarters of the adoption of the policy. Next, we document an increase in agricultural productivity following the adoption of RoW policy. The two results indicate that the RoW policies lead to an increase in the number of 4G towers as well as agricultural productivity. Furthermore, the introduction of 4G due to RoW policies is associated with increased fertilizer consumption, a convergence towards optimal N:P:K ratios, and a higher uptake of credit.

Our results are robust across various measures of agricultural productivity, employing dynamic two-way fixed effects estimation, alternative estimators suggested by Borusyak, Jaravel and Spiess (2022) and de Chaisemartin and D'Haultfuille (2020), and a falsification test employing a sub-sample of non-cropland hexagons.

Next, we investigate the improvement in access to information as a mechanism driving the increase in agricultural productivity following the introduction of 4G technology. While the introduction of 4G improves access to a wide variety of information, we focus on weather information for two reasons. First, agricultural output and returns heavily depend on weather shocks, particularly rainfall shocks (Townsend (1994), Jayachandran (2006), Kaur (2019), Rosenzweig and Udry (2013)). Second, the availability of data on weather information and its spatial distribution allows us to more credibly identify the underlying mechanism.

We start by assessing the importance of the accuracy of weather information by using the distance of a hexagon to the nearest official weather station as a proxy for the accuracy of forecasts. Almost all of the weather information and forecasts in India are sourced from the instruments located at the official weather stations. However, weather shocks, especially rainfall, exhibit low geospatial correlation (Cole and Xiong (2017)). Therefore, the distance of a hexagon to the nearest official weather station provides a proxy for the accuracy of weather information and forecasts. Specifically, as the distance to the nearest official weather station increases, the accuracy of the forecast decreases, thereby decreasing the accuracy or reliability of the information. We find that hexagons closest to the weather stations exhibit the highest treatment effect of increased agricultural productivity after the introduction of high-speed 4G internet compared to the statistically insignificant and economically negligible effects for hexagonal units far from the weather stations.

Furthermore, we refine our understanding of the mechanism by examining the heterogeneity in the treatment effects due to the potential value of information. We use historical rainfall volatility as a proxy for the value of information. The intuition behind this test is that farmers operating in regions of high rainfall uncertainty stand to gain the most from rainfall forecasts and, thus, value reliable weather information the most. Our results indicate that hexagons located closer to weather stations that are prone to greater rainfall variability are more likely to experience gains in agricultural productivity due to the introduction of 4G towers. However, the introduction of 4G towers has a small post-treatment effect for hexagons that experience low rainfall volatility.

We further supplement our results on the value of information by examining the heterogeneity in the treatment effects due to exposure to predictable and difficult-to-predict weather shocks. The reasoning behind this assertion is that weather information from a nearby weather station will be of little value to farmers in regions prone to weather shocks that are difficult to predict as forecasts will likely be less accurate. Our results indicate that treatment effects are positive, statistically significant, and economically meaningful for units that are located close to the weather stations and are susceptible to predictable shocks.

These results have two key takeaways. First, it sheds light on the mechanism, i.e., the improvement in agricultural production due to the introduction of 4G towers is related to improved access to information. Second, the introduction of infrastructure to support high-speed internet in isolation might not be adequate, i.e., it acts as a complement to existing technology. The positive effect of such infrastructure development that improves access to information crucially relies on the reliability and value of the underlying information.

Next, we examine the importance of 4G technology relative to earlier versions of telecommunications technology, such as 2G and 3G. High-speed internet in the form of 4G presents several key advantages over traditional modes of information dissemination including television, radio, phone calls, texts, and slow-speed internet. For instance, Fabregas, Kremer and Schilbach (2019) argue in their review article that internet-based platforms may be an effective tool for disseminating information in a user-friendly and engaging manner by employing videos and graphics. Moreover, 4G significantly enhances real-time and fast information accessibility for farmers through high-speed internet which is crucial for the impact of information dissemination on agricultural productivity (Anderson and Feder (2004)). We examine this conjecture by comparing the effects of the introduction of 3G and 4G on agricultural productivity. We do not find any economically or statistically significant improvement in agricultural productivity following the introduction of 3G, a generation that provides access to low-speed internet. Similarly, we do not find evidence of a positive effect of the introduction of 2G on agricultural productivity. Overall, these results suggest that 4G technology may possess distinct advantages in improving information access for farmers.

We investigate a particular channel that potentially makes information access through high-speed 4G internet special – role of trust. Information dissemination through traditional modes such as text messages, call centres, hotlines, etc. is often directly controlled by the state and its agents. Moreover, local governing bodies known as village panchayats often act as the central node for information dissemination by the government. Therefore, a lack of trust in local and state authorities can hinder effective communication and engagement through these traditional methods (Fabregas, Kremer and Schilbach (2019)). For instance, Cole and Sharma (2017) document that nearly 70% of Indian farmers distrust information provided by extension workers. On the contrary, the internet, being a more decentralized platform, may not encounter the same issues related to trust. For instance, a user of a weather application may not realize that the weather information on the app is sourced from government weather stations. Consequently, the introduction of 4G internet may have greater effects in regions where trust in state institutions is diminished.

Using India Human Development Survey (IHDS) responses to questions on the level of trust in local village panchayats and state governments, we compare changes in agricultural productivity following 4G introduction in areas with pre-treatment high and low trust levels. Our findings suggest that regions characterized by low trust in village panchayats or the state exhibit a notable increase in agricultural production following the introduction of 4G. In contrast, the impact of introducing 4G technology is economically small and lacks statistical significance in areas with a high level of trust in local village panchayats. These results imply that the advancement of high-speed internet through 4G infrastructure can mitigate the effect of human frictions, like mistrust in local institutions and the state.

Lastly, we provide evidence on technology adoption by farmers. Most of our analysis is focused on the *availability* of 4G technology through the introduction of 4G towers. However, the uptake of 4G technology by farmers - once it is accessible through 4G towers - is of first-order importance for information-related mechanisms to drive agricultural productivity growth. Moreover, the kind of information accessed by farmers may play a significant role in shaping productivity growth. We address these questions using a proprietary dataset of geocoded and time-stamped mobile application ("app") installations from Krishify – an Indian firm which runs one of the largest social networks and commerce platforms for more than ten million farmers. Our DiD estimates indicate that app installations increase after the introduction of 4G. This result makes explicit what has been implicit in our analysis till now – availability of 4G is associated with 4G adoption, as proxied through the installation of the Krishify app.

We also examine the type of information farmers access on the internet using detailed browsing and activity data of farmers on the Krishify app. The foremost four topics collectively constitute about 70% of farmer interactions on the platform. Notably, information about government programs ranks highest, contributing to 21.3% of all interactions, followed by agricultural news at 16.6%, information about farm machinery (e.g., tractors) at 16.1%, and weather information at 14.2%. Furthermore, farmers also utilize the platform to seek advice specific to animal (9.1%) and crop (8.18%) management. Interactions related to inputs like fertilizers, seeds, finance, and other miscellaneous inputs collectively make up 11.4% of all farmer engagements on the platform. The detailed activity data of farmers suggests that farmers may be using internet-based platforms extensively for gathering information on a wide variety of agri-related activities.

Related Literature: This paper contributes to four strands of literature. First, we contribute to the literature examining the effect of communications technology on agricultural yields. This literature has primarily focused on examining the effects of face-to-face, text-based, and voice-based information dissemination mechanisms on agricultural yield. Fabregas, Kremer and Schilbach (2019) presents a detailed review of the effects of extension workers, phone calls and SMS-based information programs on agricultural yields. Most prior work does not find long-run effect of such traditional communications interventions on agricultural yields.[3] We contribute to this literature in two ways. First, we document that 4G infrastructure plays a crucial role in enhancing information access through high-speed internet. Second, we show that the dissemination of information through high-speed internet yields positive effects on yields. Lastly, we complement this literature by presenting statistically insignificant results associated with the introduction of 2G and 3G technologies – which improve voice and text-based information dissemination – on yields.

The second contribution of this paper is documenting that the dissemination of information through high-speed internet may dominate conventional approaches to improving information access as it can mitigate certain frictions associated with traditional methods. The prior literature has documented that traditional approaches to information dissemination suffer from several frictions such as lack of trust (Cole and Sharma (2017)), linguistic barriers (Gupta, Ponticelli and Tesei (2023*b*)), and timeliness (Anderson and Feder (2004), Duflo, Kremer and Robinson (2011)). Particularly, the prior literature has conjectured that a lack of trust in local and state authorities can hinder effective communication and engagement through these traditional methods (Fabregas, Kremer and Schilbach (2019)). For instance, Cole and Sharma (2017) document that nearly 70% of Indian farmers distrust information provided by extension workers. We contribute to this literature in two ways. First, we document that internet based information dissemination dominates traditional modes of information dissemination. Second, we show that internet based information dissemination through 4G alleviates human frictions associated with a lack of trust in the state.

Our paper also contributes to the literature examining the effect of infrastructure development on yields. Prior work has documented limited effects of infrastructure development on yields or production. These studies include examining the effects of the construction of

[3]Cole and Fernando (2021) find no systematic impact on yields on the treated farmers in the Indian state of Gujarat that randomly received access to a hotline for agricultural advice. Gupta, Ponticelli and Tesei (2023*b*) show that a large-scale 2G expansion in India had no significant effect on agricultural yields. Additionally, experiments by Fafchamps and Minten (2012) that sent Indian farmers price and weather information via text messages found no evidence of value addition. document similar results among Colombian farmers. Other studies documenting null results of traditional information dissemination mechanism on yields include Camacho and Conover (2010), Van Campenhout, Spielman and Lecoutere (2021), Udry et al. (2019), among others. A notable exception is Casaburi et al. (2014). They document a positive effects on yields of a text message based agricultural advice experiment among small sugarcane farmers in Kenya. However, a follow-up trial of the same intervention has no significant impact on yields.

rural roads in India (Asher and Novosad (2020)), bridges in Nicaragua (Brooks and Donovan (2020)), and kisan call centers in India (Gupta, Ponticelli and Tesei (2023*b*)). We contribute to this literature by showing that 4G infrastructure can improve information dissemination among farmers and improve yields. However, our results also suggest that the introduction of infrastructure to support high-speed internet in isolation might not be adequate, i.e., it acts as a complement to existing technology. Specifically, we show that when 4G infrastructure isn't coupled with the provision of reliable and valuable information, its potential to boost output may be constrained.

Additionally, we are related to the literature examining the effect of digital infrastructure on labor market outcomes (Akerman, Gaarder and Mogstad (2015), Hjort and Poulsen (2019), Zuo (2021)), political mobilization (Manacorda and Tesei (2020)), corruption (Andersen et al. (2011), Gonzalez (2021)), confidence in governments (Guriev, Melnikov and Zhuravskaya (2021)), financial inclusion (Aker and Mbiti (2010), Suri and Jack (2016), Suri (2017), Batista and Vicente (2023)), and credit markets (DAndrea and Limodio (2023), Gupta, Ponticelli and Tesei (2023*a*)). We differ in two ways. First, while the extant literature examines the effect of 2G and 3G technology, we focus on the effect of 4G technology. Second, we examine the effect of such a technology on agricultural yields. Focusing on yields as a real economic variable is of immense interest as around 26% of the global workforce was employed in agriculture in 2019. Furthermore, enhancing agricultural productivity is essential for alleviating global poverty and addressing increasing food demands in the context of climate change, particularly for the two billion individuals residing in smallholder farming households in the developing world. (Fabregas, Kremer and Schilbach (2019)). Therefore, our results are informative for policy makers as they highlight mechanisms under which improving internet-based information access can improve the productivity of small agricultural enterprises in a developing economy, a sector grappling with stagnation.

This paper proceeds as follows. Section 2 describes the data. Section 3 delineates the empirical strategy. Section 4 presents the baseline effect of the introduction of 4G on agricultural productivity. Section 5 documents the underlying mechanisms. Section 6 examines the importance of the introduction of 4G relative to earlier versions of telecommunications technology, such as 2G and 3G. Section 7 presents the effect of 4G introduction on the adoption of 4G specific agricultural technology. Section 8 concludes.

2 Data

This section discusses the various datasets employed in the analysis. Our primary datasets include Base Transceiver Stations (BTS) geolocation data and remote sensing data on agricultural yields, croplands, and nightlights. We use the BTS data to identify the timing of the treatment and the remote sensing data to construct our key dependent variables. We also map several other economic, demographic, and agricultural variables to this dataset. Therefore, a key contribution of this paper is creating a granular database that links telecom coverage with several economic, agricultural, and demographic variables that can be employed in future research.

2.1 BTS Data

We collect a novel dataset from the Ministry of Telecom, which provides detailed information on the geolocations of all Base Transceiver Stations (BTS) in India. A BTS serves as a communication link between user devices, such as mobile phones or computers, and the network. The dataset includes an exhaustive list of all BTSs in India, which were deployed at various points in time from 1995 to 2022.

The data contains information on precise geographic coordinates of each BTS, the associated technology (e.g., 2G/3G/4G), and the date of start of operation. In total, our dataset comprises information on 2.35 million unique BTS across 717,764 towers. We restrict our sample to the subset of 1.2 million BTS that were the first-operational-BTS-by-technology at that geolocation. Figure 1 presents the spatial evolution of 4G coverage for the years 2014, 2016, 2018, and 2020. Similarly, Appendix Figure A.1 presents the evolution of 3G coverage for the years 2014, 2016, 2018, and 2020.

2.1.1 Hexagonal Tessellation

A key challenge in working with the BTS data is deciding on the appropriate unit of analysis. Particularly, it is challenging to use conventional administrative units, such as a district or block, as the unit of analysis due to two reasons. First, the geographic coverage of BTS is typically limited around the tower. Second, BTS are activated on multiple towers within close vicinity over time. As a result, the usage of large administrative units as the unit of analysis violates the assumption of homogeneous treatment within treated units, as multiple locations within a unit receive unequal treatment over time.

We address this concern, by dividing the map of India into a grid of hexagons or hexagonal tessellation, with each hexagon representing a unit of analysis. We choose hexagonal tessellation because it is the most efficient arrangement of equal circles that fills a plane without any overlaps or gaps (Birch, Oom and Beecham (2007)). Specifically, we draw hexagons with a distance of 0.1-degree latitude between their opposite ends, which is roughly equivalent to a distance of 11 km. These hexagons can be thought of as circles with a radius of 5.5 km, with approximate area of 95 sq. km. This size is reasonable considering that the range of a 4G cell tower is typically between 3-6.5 km (2-4 miles).[4] By using this hexagonal tessellation, we ensure that each location within our analysis unit receives nearly equal treatment.

We overlay the geographic coordinates of the BTS onto the tessellated map to assign each first-operational-BTS-by-technology to a unique hexagon. Appendix Figures A.2 and A.3 present the tessellated map of India showing all hexagons included in our sample. Figure 2 plots the number of hexagons first treated by BTS of different technologies from 1995 until 2022. We identify the date of treatment for each hexagon, i.e., the date on which a hexagon first saw the introduction of a 4G BTS, by using the date of operation for 4G BTS. Figure 3 presents the evolution of the treatment across hexagons over time for 4G BTSs. Similarly, Appendix Figure A.4 presents the evolution of the treatment across hexagons over time for 3G BTSs.

2.2 Remote-sensing data on Agricultural Production

We use remote-sensing data to construct measures of unit-level (hexagon-level) agricultural production, as no such granular agricultural production data exists for India. This data comes from NASA's Earth Observation Satellite - Landsat 8. Specifically, we use Enhanced Vegetation Index (EVI) to construct our unit-level measures of agricultural production.

EVI is a chlorophyll-sensitive composite measure of plant matter generated by NASA's Earth Observation satellite – Landsat 8. The composites are created from all the scenes in each 8-day period beginning from the first day of the year and continuing to the 360th day of the year. Each pixel value is optimized considering cloud cover obstruction, the influence of background vegetation, image quality, and viewing geometry. We direct the readers to Huete et al. (2002) for details on the construction of an earlier version of EVI and the usage of this earlier version – 16-day EVI composites – in economics research by Asher and Novosad (2020).

Our sample period extends from 2013 until 2021 as the 8-day EVI composites are available from April of 2013 until January of 2022. We query EVI values over this period for our desired

[4] See here.

micro-regions, hexagonal units, by supplying the geometry of the micro-region to the Google Earth Engines API. EVI values obtained from each pixel are spatially averaged over the hexagon to obtain a time-series of EVI values with an 8-day interval. Appendix Figure A.5a plots the evolution of average, the 25th percentile, and the 75th percentile EVI value for *all* hexagons in our sample.

Following the methodology outlined in Asher and Novosad (2020) and Ghosh and Vats (2022), we construct hexagon-level measures of agricultural yields for the primary growing season of kharif that begins in June and ends in October of each year. We construct two measures using the 8-day composite based EVI values. Our preferred measure – EVI implied agricultural yield – is constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum EVI value during the season. Appendix Figure A.5a shows that the EVI value increases with the start of the kharif season and reaches its highest value around the kharif harvest time. Therefore, the difference measure effectively measures production per unit of pixel during the season. Moreover, the difference measure implicitly controls for differences in non-crop vegetation, such as perennial non-crop green cover, that maybe captured by EVI composites. Our second measure – maximum EVI – uses the maximum EVI value for the kharif season. Panel A of Table 1 presents the summary statistics for the two measures. The two measures are highly correlated with a correlation coefficient of 0.53 (see appendix Figure A.5b).

An assumption that justifies the usage of these measures is that they exhibit significant correlation with both agricultural productivity measures and real production measures. We verify this assumption in Appendix B for our preferred measure of agricultural production based on 8-day composite exhibits. We find that the EVI implied yield based on 8-day EVI composites exhibit significant correlation with real production measures such as yield and production value measured in USD.[5]

2.3 Other datasets

This section describes a range of other datasets used in our analysis.

Other Remote Sensing Data: We use two other remote-sensing datasets – cropland extent and nightlights. Cropland extent data comes from the Global Food Security-Support Analysis Data (GFSAD) and identifies cropland at the resolution of 1 km for the year 2010.[6] We map the

[5] Asher and Novosad (2020) note that the agricultural production measures based on 18-day composite measures also exhibit significant correlation with both agricultural productivity measures and real production measures.

[6] GFSAD is a NASA funded project providing high-resolution global cropland data. Data documentation of can be found here.

cropland extent data to our hexagons and remove all hexagons that do not have any cropland to focus attention on agrarian units.

Weather Station Data: Data on geo-locations of weather stations and the daily weather records at these stations comes from the Indian Meteorological Department. The weather records in this data include daily rainfall and daily incidence of variety of other weather conditions, such as haze, sand/dust storm, fog, squall, gale, thunderstorm, hailstorm, fog, squall, frost, dew, and snow or sleet. This data spans from 2001 until 2019. We use this data until 2012 to construct weather related variables before the onset of the treatment.

Agricultural Production Potential: Agricultural production potential is measured using two measures provided in Advancing Research on Nutrition and Agricultures (ARENA) Demographic and Health Surveys (DHS)-GIS Database. Our measures of agricultural production from the ARENA-DHA database is based on exogenous characteristics such as soil characteristics and water availability in the area. The first measure provides the combined suitability of currently available land for pasture and rainfed crops. The second measure provides the combined suitability of the global land area for pasture and rainfed crops. Survey clusters in the DHS data are mapped to hexagons based on the minimum distance between centroids of hexagons and geo-coordinates of survey clusters. These measures are calculated as of 2014.

Trust in Village Panchayats & State Government: The data on trust in village panchayat and state government comes for the India Human Development Survey (IHDS-II) conducted in 2011, which includes a question that queries respondents about the level of trust they have in their local village panchayats and state government. The exact questions are: "How much trust do you place in the ability of your village panchayat to implement public projects?" and "How much trust do you place in the ability of the state government to take care of people?" The response options consist of three choices: a) A great deal of confidence, b) Only some confidence, and c) Hardly any confidence at all. To quantify these responses, we assign a numerical value of 1 to option (a), 0.33 to option (b), and 0 to option (c). We take the average of this measure across all households within a district based on their weights. This allows us to compute a continuous measure of confidence in village panchayats and state government at the district level. We use the response to this question to construct a district-level proxy for confidence in village panchayats and divide the districts into two subsets based on the median response value. Districts with confidence levels above the median are categorized as "high confidence" districts, while those below the median are classified as "low confidence" districts. We then link this measure to the hexagonal grids situated within the districts.

Fertilizer Consumption data: We collect district-level data on consumption of total fertilizers, fertilizer by nutrient type, and gross sown area from the States of India (SoIdx) database maintained by the Center for Monitoring the Indian Economy (CMIE). This data spans from 2000 until 2020. We map hexagons in our dataset to districts based on the extent of overlap, i.e., if a hexagon falls in multiple districts it is mapped to the district with the maximum area of overlap.

Agricultural Credit data: We obtain quarterly branch level data from one of the largest state owned bank in India. The data spans 2013-2021 and contains information about total agricultural credit disbursed by the branch during the period.

Krishify app installations data: Krishify is an Indian company aimed at connecting farmers on a social network where they can discuss agriculture related issues.[7] We obtain a proprietary, geolocated and time-stamped dataset of their app installations for the years 2019-2021. We superimpose the coordinates of app installations over our hexagonal grid to identify the number of app downloads at the hexagonal level. Additionally, we also get a 10% random sample of all farmers in the Krishify database along with their detailed search, browsing, likes, and comment history.

Table 1 presents the summary statistics of the key variables used in the analysis. The average value of our preferred measure of agricultural productivity, EVI implied yield, is 0.27 with a standard deviation of 0.14. The hexagons, on average, are located 40km away from the nearest weather station. Rainfall exhibits significant variability, as indicated by the coefficient of variation, which ranges from 0.127 to 1.414.

3 Empirical Strategy

Our objective is to estimate the effect of high-speed internet on agricultural productivity. We estimate this effect by employing a difference-in-differences (DiD) framework that leverages the staggered rollout of Broadband Transmission Stations (BTSs) within hexagons. Particularly, our empirical strategy examines the effect on agricultural productivity in hexagons around the time when BTSs first become operational in the *treated* hexagons relative to the agricultural productivity in *control* hexagons which are *not-yet-treated*.

[7] Link to their official website can be found here. Appendix section D.1 presents more details about the company and our data.

3.1 TWFE Dynamic Specification

We begin by estimating a dynamic specification 1 in relative time, which allows for non-parametric changes in treatment effects over time. Borusyak, Jaravel and Spiess (2022) argue that we need to exclude at least two relative period indicators when there are no never treated units with a panel balanced in calendar time.[8] Therefore, we exclude two relative time periods corresponding to $t \in \{-8, -1\}$ to address potential multi-collinearity concerns.

$$Y_{i,t}^{EVI} = \alpha_i + \lambda_t + \sum_{\substack{y=-8, \\ y \neq -1,-8}}^{y=8} \beta_y \mathbb{1}(t - t_i^* = y) + \varepsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ denotes the EVI-implied agricultural yield for the Kharif season in year (t) and hexagon (i). Indicator variables $\mathbb{1}(t - t_i^* = y)$ measure time relative to the year hexagon i first got treated (t_i^*). α_i and λ_t denote hexagon and time fixed effects, respectively. Standard errors are clustered at the district level.

Appendix Figure C.1 presents the sequence of dynamic coefficients $\{\beta_y\}$ estimated using equation 1. These estimates can be interpreted as causal under the assumptions of no-anticipation of the treatment, treatment effect homogeneity, and parallel trends. Negative and statistically significant coefficients for relative time periods $t = -3$ and $t = -4$ suggest meaningful pre-trends and thus, a possible violation of parallel trends assumption. However, Sun and Abraham (2021) show that testing for parallel trends using pre-treatment leads from a dynamic TWFE specification can be problematic as pre-treatment leads can be contaminated with treatment effects from other relative time periods, including post-treatment lagged effects. Therefore, a rejection of parallel trends based on significant pre-treatment leads necessarily assumes treatment effect homogeneity.

3.1.1 Treatment effect homogeneity

We begin by evaluating the plausibility of the assumption of treatment effect homogeneity in our context. Treatment effect homogeneity imposes the condition of homogeneous treatment effects across cohorts, i.e., farmers gaining access to 4G services in 2014 experience similar productivity gains as farmers that gain 4G access in 2021. Intuitively, the assumption of treatment effect homogeneity may be violated due to a variety of reasons, such as changes

[8]There are two sources of multi-collinearity. First, relative period indicators sum to one for every unit. Second, the presence of a linear relationship between two-way fixed effects and the relative period indicators.

in forecasting skills over time, mergers and acquisitions among telecom service providers, and variation across cohorts due to calendar time-varying macroeconomic conditions could govern the effects on outcomes across cohorts.

We formally diagnose the source and extent of contamination of the pre-treatment leads in our dynamic TWFE specification using the methodology outlined in Sun and Abraham (2021). The estimate for β_{-3} can be decomposed as follows:

$$\beta_{-3} = \sum_{e=2013}^{2021} \omega_{e,-3}^{-3} CATT_{e,-3} + \sum_{l \in \{-7,-6,\ldots,0,\ldots,6,7\}} \sum_{e=2013}^{2021} \omega_{e,l}^{-3} CATT_{e,l} + \sum_{l' \in \{-8,-1\}} \sum_{e=2013}^{2021} \omega_{e,l'}^{-3} CATT_{e,l'} \quad (2)$$

where $CATT_{e,l}$ refers to average treatment effect for cohort e in relative time period l. For instance, $CATT_{2015,-2}$ is the average treatment effect for the relative time period $t = -2$ for the cohort that got treated in the year 2015. The first term captures the true parameter of interest associated with $t = -3$. The third term reduces to zero under the no anticipation assumption. The second term reduces to zero under the assumption of treatment effect homogeneity. Sun and Abraham (2021) show that if treatment effects are heterogeneous, dynamic effects of 4G introduction across different cohorts will affect the TWFE estimates of β_{-3} making it a function of post-treatment $CATT_{e,l\geq 0}$. Therefore, a test of pre-trends using the statistical significance of pre-treatment coefficients becomes invalid even in the absence of pre-trends.

Appendix Figures C.2a and C.2b respectively illustrate the decomposition of weights, as described in equation 2, for the pre-treatment leads at $t = -3$ and $t = -4$ in specification 1. Weights associated with $CATT_{2015,1}$ and $CATT_{2017,1}$ receive the highest absolute weights of magnitudes of 0.079 and −0.056, respectively. If treatment effects for relative time period $t = 1$ are different across these cohorts, the interaction of non-zero weights and heterogeneous treatment effects for these cohorts will contaminate the estimates of β_{-3}. Thus, under treatment effect heterogeneity, β_{-3} becomes a function of post-treatment $CATT_{e,l\geq 0}$. making any test for pre-trends using the statistical significance of pre-treatment coefficients invalid.

3.2 DiD estimator

We address the issue of violation of the treatment effect homogeneity assumption by presenting our event-time DiD estimates using the Callaway and Sant'Anna (2021) estimator We choose this estimator for two reasons discussed in Roth et al. (2023). First, this approach provides sensible estimands even under arbitrary heterogeneity of treatment effects. Second, this approach makes transparent which units are being used as a control group to infer the unobserved potential outcomes. Specifically, we use the set of not-yet-treated cohorts as the control group because the no-never-treated group is minuscule in our design. For robustness, we also present our event-time DiD estimates using the estimators outlined in Borusyak, Jaravel and Spiess (2022) and De Chaisemartin and d'Haultfoeuille (2020).

The Callaway and Sant'Anna (2021) framework is based on causal parameters referred to as *group-time average treatment effects*. These effects represent the average treatment effect for units (hexagons) that belong to a specific group g at a particular time period t, i.e., group of hexagons treated in 2014, 2015,... etc.:

$$ATT(g,t) = \mathbb{E}[Y_t(g) - Y_t(0)|G_g = 1] \tag{3}$$

where $Y_t(g)$ is the outcome variable for treated group g at time t, $Y_t(0)$ is the untreated potential outcome for group g at time t, and G_g is a binary variable that equals one for units belonging to group g, i.e., group first treated in period g.

Most importantly, the construction of these $ATT(g,t)$'s does not rely on the assumption of treatment effect homogeneity across groups or across time. These $ATT(g,t)$'s are then aggregated across groups to arrive at event-time parameters analogous to coefficients in a dynamic TWFE specification as in equation 1 while avoiding the pitfalls associated with such specifications.

3.2.1 Identifying Assumption

Our empirical strategy exploits the variation across treated and not-yet-treated hexagonal units (hexagons). Therefore, one natural concern is that early treated units are different from units that are treated later. However, as long as these differences are time-invariant, this selection is fully accounted for by unit fixed effects. Thus, our specification does not require that the introduction of high-speed 4G BTS was randomly allocated, nor does it require that units must have the same pre-treatment characteristics. Rather our estimate would be unbiased if the treated units would have evolved the same way as the not-yet-treated units in the absence of the treatment.

Our identifying assumption implies that unit-level trends in agricultural production would have been the same in treated and not-yet-treated in the absence of the policy. While this assumption is untestable, we present parallel pre-trends for all our analysis to provide supporting evidence to our identifying assumption.

4 Results

This section presents the effect of 4G introduction on agricultural yields. We use the heterogeneity-robust, dynamic difference-in-differences estimator outlined in Callaway and Sant'Anna (2021) that exploits the staggered introduction of 4G across hexagons in our sample.

Figure 4 presents the results using the Callaway and Sant'Anna (2021) estimator.[9] The dependent variable is standardized to mean zero and standard deviation of one. Our sample consists of all hexagons that received the 4G treatment at some point between 2013 and 2021, thus comparing treated units with not-yet-treated units. Our sample is a balanced panel of hexagons from 2013 until 2021. We restrict our sample to agrarian units by removing hexagons that did not contain any cropland within its boundaries as identified by the GFSAD data on cropland extent. The outcome variable is the EVI implied yield, constructed by subtracting the average value of EVI during the initial weeks of the kharif season from the maximum EVI value over the entire season. Each plotted coefficient measures the average treatment effect on the treated in event-time e constructed by aggregating over event-time coefficients for groups treated in each year, i.e., $ATT(g, g+e)$'s.

There are two key takeaways from Figure 4. First, we do not observe significant pre-trends in EVI implied yields in the event-time leading up to the treatment. Second, we observe an increase in EVI implied yield following the introduction of high-speed 4G internet. Specifically, we observe that the treatment effect increases gradually after treatment. We find that six years after the treatment EVI implied yield increases by 0.24 times the standard deviation or 0.035 unit increase in EVI. This increase is equivalent to the 12.5% of the mean value of EVI implied yield.

4.1 Discussion of the Magnitude of Estimate

We translate the estimates of 4G introduction on EVI implied yield to the effect on real agricultural yield and production value. Appendix B presents the detailed calculations. Our estimates imply that the introduction of high-speed 4G internet is associated with a – (1) 1.06% increase in agricultural yields of cereal crops, (2) a 4.28% increase in the production value of cereal crops, and (3) a 3.21% increase in the production value of all crops. Overall, we find that six years after the introduction of 4G internet, the annual income of agricultural house-

[9] As pointed out in Roth (2024), the default pre-treatment plots for the Callaway and Sant'Anna (2021) estimator in Stata do not have the same interpretation as those of the traditional TWFE event-study plots. We correct for this difference by including the `long2` option in our Stata specification for the Callaway and Sant'Anna (2021) estimator. The results are presented in Figure C.3.

holds grew by 14.5%.[10] According to the Situation Assessment Survey of India, the income of agricultural households grew by 59.01% over the period from 2014 until 2021. Our simple calculations suggest that 24.56% of this income growth of agricultural households over this period can be explained by the introduction of high-speed 4G internet.

We compare our estimate of the increase in yield with the effects of climate change, infrastructure developments, and other interventions on agricultural yields. Guiteras (2009) estimates that projected climate change in India will reduce major crop yields by 4.5-9% over the period 2010-2039. Our estimate of increase in yield of 1.02% for major cereal crops, six years after the introduction of high-speed 4G internet, may be sufficient to offset this projected climate change impact on agricultural yield in India.

Next, we compare our estimates to the documented effects of other infrastructure development on agricultural yields. Asher and Novosad (2020) study the development of rural road infrastructure in India, and do not find any significant effect on agricultural income four years after road construction. Similarly, Brooks and Donovan (2020) do not find a statistically significant effect of construction of bridges in Nicaragua on agricultural production of maize and beans. Gupta, Ponticelli and Tesei (2023*b*) examine the effect of introduction of kisan call centers in India and do not find any significant effect on agricultural yields, six years after their introduction. In contrast, our estimate indicates a 1.02% increase in agricultural yield, five years after the introduction of high-speed 4G internet.

Our estimate is smaller when compared to other agricultural interventions that reduce risk or directly increase productivity. Emerick et al. (2016) conduct a randomized controlled trial in India that involves distribution of a new rice variety that reduces downside risk by providing flood tolerance. They document that this intervention has positive effects on adoption of a more labor-intensive planting methods, area under cultivation, fertilizer usage, and credit utilization translating to improvement in productivity. Their most conservative estimate indicates that productivity increases by 6%. Our estimate of increase in yield is roughly one-sixth of the estimate reported in Emerick et al. (2016). Beaman et al. (2013) document that yield increases by 31% and 16.6% when farmers in Mali receive fertilizers equal to the recommended quantity and half the recommended quantity, respectively.

[10]We direct readers to Appendix B.1 for the detailed calculations.

4.2 Overall Effects of Treatment

We also provide estimates of the overall effects of treatment. We characterize the overall treatment effect using three different estimators outlined in Callaway and Sant'Anna (2021) that aggregate $ATT(g,t)$'s into a single estimate of the aggregate average treatment effect. First, in simple aggregation, the aggregate ATT is defined as the weighted average of all group-time average treatment effects. Second, in group-specific aggregation, the aggregate ATT is defined as the average of partially aggregated group-specific ATT's across all groups. Third, in event-time aggregation, the aggregate ATT is defined as the average of partially aggregated event-time ATT's across all lengths of post-treatment exposure.

Table 2 presents the three measures of aggregate ATT. The last column of Table 2 presents our estimates of the average effect of 4G introduction on agricultural yields. The estimates indicate that the introduction of high-speed 4G internet significantly increased agricultural yields across all three measures of aggregate ATT.

4.3 Robustness

This section investigates the validity of our empirical design and the robustness of our primary results. We conduct a falsification and a sanity test that exploits heterogeneity in the treatment effect by agricultural production potential to provide confidence in our empirical design and our hypothesized effect. Lastly, we present the robustness of our results across two dimensions – an alternative measure of agricultural production and alternative estimation methodologies.

Falsification Test: First, we present a falsification test using a sub-sample of hexagons that do not contain any cropland within its boundaries as of 2010. The objective of this test is to show that our empirical design is valid, the results are unlikely to be driven by spurious factors, and our results represent an actual increase in agricultural production. The falsification sub-sample contains no cropland. Therefore, we should not observe any change in agricultural production in these units. Appendix Figure C.4 presents these results. We do not find any economically or statistically significant effect on agricultural production for these units.

Heterogeneity by Agricultural Production Potential: Second, we present a sanity check for our results by exploiting the heterogeneity in agricultural production potential. Intuitively, if high-speed internet increases agricultural production, we should document a greater effect in areas with high potential for agricultural production. Appendix Figure C.5 presents these results for two measures of agricultural production measured as of 2014. Our measure of agricultural production from the ARENA-DHA database is based on exogenous characteristics such as soil characteristics and water availability, rather than actual agricultural production. We find that the effect of the introduction of high-speed 4G internet is higher in areas with high agricultural production potential (shown in red) relative to areas with low agricultural production potential (shown in blue). Quantitatively, the effect in areas with high agricultural production potential is twice as high as the effect in areas with low agricultural production potential.

Alternative Measure: Third, we demonstrate the robustness of our results using an alternative measure of agricultural production. We use the maximum EVI value during the kharif season as our dependent variable. Appendix Figure C.6 presents these results and documents results similar to our baseline results.

Alternative Estimation Methodology: Fourth, we test the robustness of our results using alternative estimators proposed in Borusyak, Jaravel and Spiess (2022) and de Chaisemartin and D'Haultfuille (2020).[11] Figure 5 presents the dynamic event-time coefficients obtained from these estimators, with our baseline measure of agricultural production — EVI implied yield – as the dependent variable. For reference, we also show the dynamic TWFE estimator and our baseline estimator of Callaway and Sant'Anna (2021). Lastly, we repeat this exercise using maximum EVI as the dependent variable, shown in Appendix Figure C.7. The tests employing alternative estimation methodologies echo our baseline results.[12]

4.4 Alternative Identification Strategy

So far, our estimation strategy has relied on the staggered roll-out of 4G technology throughout the country. While our identification strategy does not require that the early and later treated units have the same pre-treatment characteristics, there may still be concerns associated with the endogeneity in the spatial introduction of 4G technology. We address such concerns by exploiting plausibly exogenous variation in the roll-out of 4G technology due to the staggered adoption of a policy meant to promote the growth of telecom infrastructure at the state level. Specifically, we exploit the staggered adoption of Rights of Way (RoW) policies by different states between 2015 and 2019 to capture plausibly exogenous variation in the establishment of 4G telecom towers.

[11] We do not use the Sun and Abraham (2021) estimator as it uses the set of never treated units as the control group, which is very small in our setting. Therefore, the Sun and Abraham (2021) estimator is inappropriate in our setting.

[12] The standard errors in the pre-period for the Borusyak, Jaravel and Spiess (2022) estimator are large. This issue of the large standard errors is consistent with the discussion on standard error estimation in Borusyak, Jaravel and Spiess (2022) and Braghieri, Levy and Makarin (2022). Similar to Braghieri, Levy and Makarin (2022), we also find that the standard errors increase dramatically as we increase the number of pre-periods.

4.4.1 Details of the RoW policy

Laying of optical fibers and installing telecom towers requires significant digging and trenching work on public lands (GSMA (2020)). As such, access to public Rights of Way ("RoW") forms an essential part of deploying telecom infrastructure at a large scale. The permissions required for such RoW are the jurisdiction of states and local municipalities.

However, the telecom industry often faced strong opposition from local bodies and the public due to issues related to estate taxes and concerns over electromagnetic frequency emissions that allegedly cause cancer. For instance, the Cellular Operators Association of India (COAI), the key lobby group for the major telecom operators in the country, presented its grievances to the telecom regulator, saying, "State bodies continue to initiate actions such as disconnecting electricity, sealing the premises and even dismantling of tower sites without any prior notice, leading to coverage disruptions and network congestion."[13]

In order to standardize rules across the country and reduce the regulatory burden on telecom infrastructure companies, the federal government laid down a framework to regulate the establishment and maintenance of telecom infrastructure to coordinate policies at the national level. This effort led to the introduction of the Indian Telegraph Right of Ways Rules on 15th November 2016. The policy provides a blueprint for rules to be followed by telecom companies and local administrative bodies. Anecdotal evidence suggests that the blueprint was inspired by the RoW policy implemented in the state of Jharkhand in December 2015 following a joint pilot between the Telecom Regulatory Authority of India (TRAI) and the government of Jharkhand.[14]

These rules were designed to streamline the process by which companies apply for approvals and resolve disputes related to telecom infrastructure. Moreover, the rules required the establishment of specific timeframes. For example, the policy dictates that the state authority must respond to applications within 60 days from the date of submission, either by granting or denying the application. If the state authority fails to respond within this timeframe, permission is automatically granted, and the authority cannot reject the application without first giving the company an opportunity to be heard and providing a written reason for the rejection.

While the policy was drafted by the federal government as a blueprint, its adoption was left to the discretion of individual states. Consequently, different states implemented the RoW framework at various times between 2015 and 2019. Appendix Table C.2 presents the specific

[13] See here for more details.
[14] See here for details.

adoption dates of the RoW framework by different states. Sixteen states had adopted the framework by December 2019.

The policy was perceived as a positive development by the telecom industry. For instance, Rajan Matthews, director general of the COAI, noted, "It is a great move to assist the industry with improving the quality of service experience of customers...This will provide a great fillip to expanding cell site coverage as well as fiber implementation to support broadband services."[15] Overall, the passage of a state-wide RoW policy significantly reduced the regulatory and operational burden on telecom companies.

4.4.2 Identifying Assumption

The selection of states into the adoption of RoW policies could be endogenously linked to agricultural productivity.[16] We address this concern in two ways. First, we use the set of adopters, i.e., states that eventually adopted RoW policy by 2020. Second, we include unit and time-fixed effects in our empirical specification. As before, we use a heterogeneity-robust Callaway and Sant'Anna (2021) estimator. Therefore, our empirical strategy exploits variation across treated and not-yet-treated states. Moreover, our estimation strategy fully accounts for selection concerns as long as the differences across states that lead to the adoption of RoW policies are time-invariant. Under this identifying assumption, the date of adoption of RoW policies is likely to be random or unlikely to be driven by pre-existing differences across adopters.

4.4.3 First Stage – Effect of RoW Adoption on Installation of 4G Telecom Towers

We quantitatively establish the effect of the staggered adoption of the RoW rules across states on the installation of telecom towers. Figure 6 plots the dynamic event-time coefficients using the Callaway and Sant'Anna (2021) estimator. The unit of analysis is at the district level. The outcome variable is the number of operational 4G BTSs in districts measured at the quarterly frequency from 2013 Q1 to 2021 Q4.[17] For robustness, we report results without any controls as well as after controlling for the natural logarithm of the area of the district. Our results indicate that the adoption of RoW policy lead to a significant increase in the number of 4G

[15]See here for more details.

[16]For instance, states with a proactive executive and bureaucratic branch are more likely to be early adopters of a high-impact policy like RoW policy. The proactive nature of the state machinery might have effects in other policy domains related to agriculture and, thus, be correlated with agricultural productivity at the state level.

[17]We exploit the granular information on set up of 4G BTSs in our dataset to move our analysis a step finer, up to the quarterly frequency. The annual frequency of our analysis so far has been dictated by the annual frequency of the EVI-implied agricultural yield.

tower installations. Specifically, the number of towers within a district increased by 600 within six quarters of the adoption of the policy, and the effect tapers off thereafter. The increase in the number of operational 4G towers after the adoption of RoW policy is consistent with the anecdotal evidence that the RoW policy significantly reduced the regulatory and operational burden on telecom companies.

4.4.4 Effect of RoW Adoption on Agricultural Production

Next, we examine the effect of the adoption of the RoW policy on agricultural production. Figure 7 plots the dynamic event-study coefficients using our preferred Callaway and Sant'Anna (2021) estimator. There are two key takeaways from Figure 7. First, we do not observe significant pre-trends in EVI implied yields in the event time leading to the treatment. Second, we observe an increase in EVI implied yield following the adoption of RoW policy. Specifically, we observe that agricultural production increases by 0.8 times the standard deviation or 0.11 unit increase in EVI implied yield two years after the adoption of the RoW policy. This increase is equivalent to the 39.1% of the mean value of EVI implied yield. For robustness, we also supplement Figure 7 with OLS TWFE estimator, Borusyak, Jaravel and Spiess (2022) estimator, and de Chaisemartin and D'Haultfuille (2020) estimator. Overall, our results using the alternative identification strategy of staggered adoption of RoW policy indicate that our baseline results are unlikely to be driven by spatial endogeneity associated with the installation of 4G towers.

4.4.5 Effect on Operational Inputs – Fertilizers

This section documents that the relationship between telecom entry on usage of agricultural inputs such as fertilizers. The intuition behind this analysis is that the dissemination of information through 4G can decrease uncertainty, ultimately encouraging higher levels of investment (Bloom (2009)). Specifically, the introduction of 4G towers, which increases access to information about upcoming weather shocks, is expected to lead to higher usage of inputs as farmers are more well-equipped to navigate these shocks (Rosenzweig and Udry (2013)).

We use the annual district-level data on fertilizer usage to test the association between 4G penetration and fertilizer consumption. Specifically, fertilizer consumption is measured as the natural logarithm of the amount of consumption of total fertilizers (NPK), nitrogen (N), phosphate (P), and potash (K) fertilizers per unit of gross sown area. We map our hexagons to districts and construct the fraction of hexagons with at least one 4G tower in that year as a district-level measure of 4G penetration. Appendix Figure C.8 presents the local-polynomial plot between 4G penetration and fertilizer consumption. The figure presents preliminary evidence indicating a positive relationship between fertilizer consumption and 4G penetration.

We further refine the analysis using a regression setup that includes district and year fixed effects. Appendix Table C.1 presents the results. Consistent with the preliminary evidence presented in appendix Figure C.8, we find a positive association between 4G penetration and fertilizer consumption. Results in column (1) indicate a 10% increase in 4G penetration is associated with a 1.8% increase in total fertilizer consumption per unit of gross sown area. Similarly, results in column (2), (3), and (4) indicate that an increase in 4G penetration is associated with an increase in consumption of nitrogen, phosphorus, and potassium fertilizer consumption per unit of gross sown area, respectively.

We further strengthen our argument by examining the effect of staggered adoption of Rights of Way (RoW) rules across states on fertilizer consumption. Figure 8 presents the results. We document an increase in the total fertilizer consumption after the adoption of Rights of Way rules that lead to the rapid introduction of 4G towers. Moreover, the increase in fertilizer consumption does not seem to be driven by pre-trends. Specifically, we find that the total fertilizer consumption is driven by an increase in the consumption of phosphate and potash based fertilizers. Meanwhile, the consumption of nitrogen based fertilizers does not increase to a similar extent. Specifically, we find that the total fertilizer consumption per unit of gross sown area increased by 21.88% after the adoption of Rights of Way rules. While the consumption of nitrogen based fertilizers increased by 8.6%, the consumption of phosphate and potash based fertilizers increased by 72.72% and 89.54%.

This is not surprising given that nitrogen based fertilizers are extensively used in Indian agriculture whereas the usage of potash and phosphate based fertilizers is very limited. For instance, Panel C of Table 1 indicates that on average 101,102 kg of Nitrogen based fertilizers are used per hectare of gross sown area. In contrast, only 39,230 kg of phosphate and 14,230 kg of potash-based fertilizers are used per hectare of gross sown area. In fact, the ratio of nitrogen to phosphate to potash (N:P:K) is 7.10:2.76:1.00. This observed NPK ratio is quite different from the recommended optimal ratio of 4:2:1.

4.4.6 Effect on N:P:K Mix

An increase in the consumption of phosphate and potash based fertilizers, while keeping the nitrogen consumption fixed, could be a positive outcome if the N:P:K ratio moves closer to the optimal ratio of 4:2:1. This section examines the effect of RoW rules adoption on the optimal ratio. Intuitively, if the adoption of RoW rules leads to improved 4G coverage, which in turn enhances access to information, we expect that the distance of N:P:K consumption ratio from the optimal ratio in treated areas will either converge to zero or at least decrease. To put it more directly, when farmers have better access to information about agricultural production practices, they are more likely to shift from an inefficient fertilizer mix to a more optimal one.

To this end, we compute the Euclidean distance of the consumption of N:P:K ratio from the optimal ratio of 4:2:1 using the following measure:

$$Distance = \sqrt{\left(\frac{N}{N+P+K} - \frac{4}{7}\right)^2 + \left(\frac{P}{N+P+K} - \frac{2}{7}\right)^2 + \left(\frac{K}{N+P+K} - \frac{1}{7}\right)^2} \tag{4}$$

On average, the NPK ratio deviates from the optimal ratio by 0.17, which indicates a significant difference from zero.[18] If fertilizer usage adhered perfectly to the optimal ratio rule, the value of this distance measure would be zero.

Figure 9 presents the results. We observe that the N:P:K distance, as calculated using equation 4, decreases for the treatment units compared to the control units after the adoption of RoW rules. Specifically, if the ratio is moving closer to the optimal rule, the decline should be around 0.16, as indicated by the maroon dashed line. Figure 9 suggests that the extent of decline in the relative distance for the treatment group is approximately equal to the average distance from the optimal ratio before the treatment. Furthermore, these results do not seem to be driven by pre-trends. In summary, our findings indicate that the N:P:K ratio tends to approach the optimal ratio for these essential nutrients following the adoption of RoW rules.

4.4.7 Effect on Financial Input – Credit

This section examines the impact of 4G introduction on agricultural credit using branch-level data from one of the largest public-sector banks in India. Apart from the productivity shock at the farmer level, greater awareness, better monitoring, and reduced information asymmetry fostered by the introduction of 4G can lead to greater demand and supply for agricultural credit.

[18] The t-statistic associated with a test to examine the difference of the distance measure from 0 is 158.03.

We combine the branch-level quarterly credit data with the dates of adoption of RoW policies by various states to examine the effect of 4G introduction on agricultural credit. Figure 10 presents the results using the Callaway and Sant'Anna (2021) estimator. The results suggest that agricultural credit experiences a 20% increase one year after the implementation of the Right of Way (RoW) laws and maintains a stable level thereafter.

Moreover, we aggregate the branch-level credit data at the zip code level. This aggregation at the zip code level allows us to examine the effect of the actual introduction of 4G BTS. Appendix Figure C.9 presents the results examining the impact of the staggered introduction of 4G BTS on agricultural bank credit. Our results are qualitatively similar to the results reported in Figure 10 and indicate an increase in bank credit following the introduction of 4G BTS in the zip codes.

5 Mechanism

This section investigates the mechanism driving the increase in agricultural productivity following the introduction of 4G technology. Specifically, we document the significance of enhanced access to timely information due to the availability of high-speed 4G internet connectivity. Moreover, we show that high-speed internet in isolation might not be adequate to improve output. The ultimate impact of improved information accessibility on agricultural output hinges on factors such as the reliability, accuracy, and value of the information. Therefore, infrastructure development improving high-speed internet is likely to be a complement.

For the majority of this section, we will focus on weather information as agricultural output and returns heavily depend on weather shocks, particularly rainfall shocks.[19] As such, access to information about the season's impending weather shocks could change the agricultural output. Specifically, access to the forecast of the season's weather shocks can allow farmers to choose inputs more appropriately leading to significant improvement in agricultural productivity (Rosenzweig and Udry (2013)).[20] Additionally, information about the timing and the quantity of rainfall plays a crucial role in choosing the timing of plowing, planting, cultivating, fertilizing, and harvesting crops.

[19]Prior work has documented that rainfall fluctuations can have a significant impact on aggregate consumption in village economies and agricultural productivity (Townsend (1994), Jayachandran (2006), Kaur (2019), among others).

[20]For instance, Rosenzweig and Udry (2013) find if rainfall were at the mean of the observed historical rainfall distribution in India, ₹10,000 increase in planting-stage investments would lead to an increase in profits of about ₹20,000 (over a base of ₹33,000). Similarly, if rainfall were at the 75th percentile, the same increase in plantingstage investment would generate an additional profit of ₹40,000.

5.1 Reliability of Information

We posit that the introduction of high-speed 4G internet increases agricultural productivity by improving access to weather information for farmers. We test this claim by exploiting the heterogeneity in the distance to the nearest official weather station. Specifically, as the distance to the nearest official weather station increases, the accuracy of the forecast decreases, thereby decreasing the accuracy or reliability of the information.

The distance of a hexagon to the nearest official weather station provides a proxy for the accuracy of forecasts for three reasons. First, almost all of the weather information and forecasts in India are sourced from the official weather stations maintained by the Indian Meteorological Department (IMD). IMD has been issuing annual forecasts of the monsoon across the subcontinent since 1895, and farmers appear to respond to these forecasts (Rosenzweig and Udry (2013)). Second, weather shocks, especially rainfall, exhibit low geo-spatial correlation. This fact has been widely discussed in the context of demand for weather insurance contracts, wherein the distance from the nearest weather station increases basis risk due to low geo-spatial correlation (Cole and Xiong (2017), Robles et al. (2021), Ghosh and Vats (2022)).[21] Third, farmers seem to be aware of the issues due to the distance to the nearest weather station. Using primary data from India, Cole, Giné and Vickery (2017) document that farmers view basis risk due to the distance to the nearest weather station as a significant drawback of an insurance product.

India has a network of 476 official weather stations distributed across its expanse. We map each hexagonal unit to a unique weather station that is geographically closest. On average, each station is connected to approximately 60 hexagonal units. The average distance between hexagons and their closest weather station is approximately 40 km, with a standard deviation of 19 km. Notably, there exists significant variability in the distances to the closest weather station: the 25th percentile value stands at 26 km, while the 75th percentile value reaches 52 km. We divide the hexagonal units into three distinct sub-samples based on their proximity to the nearest weather station. The first sub-sample, termed low, encompasses hexagons with distances below the 25th percentile value. The second sub-sample, medium, comprises hexagons with distances greater than or equal to the 25th percentile value but less than or equal to the 75th percentile value. Lastly, the third sub-sample, high, includes hexagons with distances above the 75th percentile value.

[21] Hill, Robles and Ceballos (2016) document that doubling the distance to a reference weather station increases basis risk and decreases insurance demand in India by 18%. Mobarak and Rosenzweig (2013) estimate that for every kilometer increase in the (perceived) distance of the weather station for a farmer without any informal risk protection, there is a drop-off in demand for formal index insurance of 6.4 percent.

Figure 11 presents the DiD estimates for the three sub-samples. As before, we do not observe significant pre-trends in EVI implied yields in the event time leading to the treatment. The most significant takeaway from this figure is the heterogeneity in the treatment effect across the three sub-samples. Hexagons closest to the weather stations exhibit the highest treatment effect of increased agricultural output after the introduction of high-speed 4G internet. Whereas hexagonal units far from the weather stations observe statistically insignificant and economically negligible increase in agricultural productivity. Specifically, the treatment effect decreases when weather information becomes less reliable, i.e., the distance to the official weather station increases.

5.2 Value of Information

We refine our understanding of the underlying mechanism by examining the heterogeneity in the treatment effects due to the potential value of information. We begin by using historical rainfall volatility as a proxy for the value of information. The intuition of this assertion is that farmers operating in regions of high rainfall uncertainty stand to gain the most from rainfall forecasts and, thus, value reliable weather information the most.

We use historical rainfall recorded at the official weather station from 2001 until 2012, one year before the start of the DID sample, to compute rainfall volatility. Specifically, we compute the coefficient of variation (CV) for monsoon rainfall and divide stations into two distinct sub-samples based on their CV value. The first sub-sample, termed low rainfall variability, includes stations with CV values below the median. The second sub-sample, termed high rainfall variability, comprises stations with CV values greater than the median. Furthermore, we split the hexagons into two sub-samples based on the distance to the nearest weather station. We consider a weather station as valuable if the history of monsoon rainfall recorded at that station exhibits high volatility, i.e., the rainfall forecasts are likely to be most valuable for hexagonal units that have a history of high rainfall volatility and are located closer to the weather stations.

Figure 12 presents the results. Panels 12a and 12b report the results for stations with low and high rainfall volatility. Results in Figure 12a indicate that the introduction of 4G towers has a small post-treatment effect for hexagons that experience low rainfall volatility. Figure 12b shows that hexagons located closer to weather stations that are prone to greater rainfall variability are more likely to experience gains in agricultural productivity due to the introduction of 4G towers. Meanwhile, we do not find statistically significant or economically meaningful treatment effects for hexagons far from the weather stations (blue diamonds in Figure 12b).

5.3 Accuracy of Forecasts

We further supplement our results on the value of information by examining the heterogeneity in the treatment effects due to exposure to predictable and difficult-to-predict weather shocks. Some regions are prone to weather shocks that are difficult to predict. Reliable weather information from a close-by weather station will be of little value to farmers in such regions as forecasts will likely be less accurate. Whereas reliable information from a close-by weather station is likely to be valuable for farmers operating in regions exposed to predictable weather shocks as forecasts are likely to be more accurate. Overall, the intuition of this test is that predictable weather shocks are likely to be associated with better forecasts. Specifically, Rosenzweig and Udry (2013) note that more accurate forecasts allow farmers to make optimal input choices ex-ante and significantly increase yields and returns for the season.

Predictable weather shocks include mist, drizzle, and rainfall.[22] We define a unit as susceptible to predictable weather shock if the probability of the event occurring, based on daily historical data of the occurrence from 2001 until 2012, is below the sample median value. We use units below the median value as the low incidence of these events is associated with decreased agricultural production. Difficult-to-predict weather shocks include haze, sand or dust storm, fog, squall, gale, and hailstorms. We define a unit to be susceptible to difficult-to-predict weather shock if the probability of the event occurring, based on daily historical data of the occurrence of these events from 2001 until 2012, is above the median value for the sample. We use units above the median value as the high incidence of these events is adversely related to agricultural production.

Figure 13 presents the results. Specifically, panels 13a and 13b present the results from the sub-sample of units with high exposure to predictable and difficult-to-predict weather shocks, respectively. Figure 13a shows that dynamic treatment effects are positive, statistically significant, and economically meaningful for units located close to the weather stations and are susceptible to predictable shocks. However, we do not find evidence of improved agricultural productivity after the introduction of 4G towers for units located farther away from weather stations. Meanwhile, Figure 13b shows that the treatment effects are statistically

[22] We want the readers to note that we are not arguing that these weather shocks can be perfectly predicted. We simply mean that these weather shocks are relatively easier to predict than difficult-to-predict weather shocks.

insignificant and economically small for units more exposed to difficult-to-predict weather shocks, regardless of their distance to weather stations.

Overall there are two key takeaways from the analysis in this section. First, the results suggest that one channel through which high-speed 4G internet increases agricultural output is through increased information accessibility. Second, in cases where the internet isn't coupled with the provision of reliable and valuable information, its potential to boost output may be constrained. Specifically, high-speed internet serves as the conduit for information, but the actual yield stems from the information. The positive effect of such infrastructure development that improves access to information crucially relies on the availability of reliable and valuable information. Therefore, the introduction of infrastructure to support high-speed internet in isolation might not be adequate, i.e., it acts as a complement to existing technology.

6 Is High-Speed Internet Special?

This section examines the importance of the introduction of 4G relative to earlier versions of telecommunications technology, such as 2G and 3G. The intuition for comparing the different technologies arises from the commonality that all have the potential to enhance access to information. However, 4G presents several key advantages over 3G and other earlier versions. First, internet access through 4G has the potential to significantly enhance real-time information accessibility for farmers. In their review article, Anderson and Feder (2004) argue that the impact of information dissemination on agricultural productivity hinges on farmers' timely access to relevant information. Second, internet-based platforms can provide valuable insights in a user-friendly and engaging manner by employing videos and graphics. Video-based interventions have been found to improve knowledge and farmers' practices (Vasilaky et al. (2015), Van Campenhout, Spielman and Lecoutere (2021)). Third, internet-based platforms do not face limitations similar to the ones faced by traditional text or call-based information delivery systems. For example, farmers might encounter challenges in reading text messages or experiencing delays in connecting with attendants at farmer call centers (Fabregas, Kremer and Schilbach (2019)). Additionally, interactions based on voice may encounter language barriers, as highlighted in Gupta, Ponticelli and Tesei (2023*b*).

We examine this conjecture by comparing the effects of the introduction of 3G and 4G on agricultural productivity. Figure 14 presents the results examining the effect of 3G introduction on agricultural productivity. Unlike the results in Figure 4, we do not find any economically or statistically significant improvement in agricultural productivity following the introduction of 3G. For completeness, appendix Figure D.1 presents the results examining the effect of 2G introduction on agricultural productivity. Similar to 3G, we do not find evidence of an economically or statistically significant improvement in agricultural productivity following the introduction of 2G.

We also replicate the results examining the introduction of 3G BTS on agricultural bank credit in Appendix Figure D.2. We note a qualitative increase in agricultural bank credit following the introduction of 3G BTS, but the results appear to be statistically insignificant. Moreover, the magnitude of the effect of 3G is smaller relative to the magnitude of the effect of the introduction of 4G BTS.

The different effects of the introduction of 2G, 3G and 4G indicate that timely information access through high-speed internet is special and is associated with improvements in agricultural productivity. The null effects observed post the introduction of 3G aligns with prior literature, which has consistently shown scant evidence supporting enhanced agricultural productivity through conventional methods of information dissemination. Anderson and Feder (2004) and Duflo, Kremer and Robinson (2011) have attributed the lack of timely and personalized information to farmers for the poor performance of traditional face-to-face programs. Casaburi et al. (2014) show that sending text messages with agricultural advice to small sugarcane farmers in Kenya didn't result in long-term positive effects on yields. In an experiment in the Indian state of Gujarat, Cole and Fernando (2021) find no systematic impact on yields on the treated farmers that randomly received access to a hotline for agricultural advice. Gupta, Ponticelli and Tesei (2023*b*) show that a large-scale 2G expansion in India had no significant effect on agricultural yields. Additionally, experiments by Fafchamps and Minten (2012) that sent Indian farmers price and weather information via text messages found no evidence of value addition. Similar results were documented by Camacho and Conover (2010) among Colombian farmers. We direct readers to Fabregas, Kremer and Schilbach (2019) for a more detailed review of the effects of extension workers, phone calls and SMS-based information programs on agricultural yields.

These results are consistent with the idea that high-speed internet is special and has the potential to activate various channels that boost agricultural productivity.

6.1 Why Might High-Speed Internet be Special?: Role of Trust

This section examines a potential channel that may make high-speed internet access, due to introduction of 4G, an effective tool for information dissemination. Specifically, we examine the heterogeneity in the treatment effect of 4G introduction by the trust in local and state authorities.

Information dissemination through traditional modes such as text messages, call centres, hotlines, etc. is often directly controlled by the state and its agents. Moreover, local governing bodies known as village panchayats often act as the central node for information dissemination by the government. Therefore, a lack of trust in local and state authorities can hinder effective communication and engagement through these traditional methods (Fabregas, Kremer and Schilbach (2019)). For instance, Cole and Sharma (2017) document that nearly 70% of Indian farmers distrust information provided by extension workers. On the contrary, the internet, being a more decentralized platform, may not encounter the same issues related to trust.[23] Consequently, the introduction of 4G internet may have greater effects in regions where trust in state institutions is diminished,

We investigate this hypothesis by combining the India Human Development Survey (IHDS) with our dataset. Specifically, we utilize data from IHDS-II, conducted in 2011, which includes a question that queries respondents about the level of trust they have in their local village panchayats.[24] We use the response to this question to construct a district-level proxy for confidence in village panchayats and divide the districts into two subsets based on the median response value. Districts with confidence levels above the median are categorized as "high confidence" districts, while those below the median are classified as "low confidence" districts. We then link this measure to the hexagonal grids situated within the districts.

Figure 15a presents the results. The outcomes for districts with low confidence levels are displayed in red, while the results for districts with high confidence levels are shown in blue. The findings suggest that regions characterized by low trust in village panchayats exhibit a notable increase in agricultural production. In contrast, the impact of introducing 4G technology is economically marginal and lacks statistical significance in areas with a high level of trust in local village panchayats.

We also present results using an alternative measure of trust, i.e., the confidence in the state government to look after people.[25] Similar to before, in regions where trust in state

[23] For instance, while looking at the Apple Weather App many may not realize that the app ultimately sources all its data from the National Weather Service, which is part of the National Oceanic and Atmospheric Administration (NOAA), a scientific and regulatory agency within the United States Department of Commerce, a United States federal government department.

[24] The exact question is as follows: "How much trust do you place in the ability of your village panchayat to implement public projects?" The response options consist of three choices: a) A great deal of confidence, b) Only some confidence, and c) Hardly any confidence at all. To quantify these responses, we assign a numerical value of 1 to option (a), 0.33 to option (b), and 0 to option (c). We take the average of this measure across all households within a district based on their weights. This allows us to compute a continuous measure of confidence in village panchayats at the district level.

[25] The two variables trust in village panchayat and trust in state government are highly correlated with each other with a

government is low, the marginal benefit of the introduction of 4G internet for information-sharing is high. Figure 15b presents the results. The findings suggest that regions characterized by low trust in state government exhibit a notable increase in agricultural production. In contrast, the impact of introducing 4G technology is economically marginal and lacks statistical significance in areas with a high level of trust in state government.

Overall these results imply that the advancement of high-speed internet through 4G infrastructure can mitigate the effect of human frictions, like mistrust in local institutions. These barriers typically hinder the dissemination of information through traditional channels.

7 Technology Adoption: Evidence using app downloads

This section examines the effect of 4G introduction on the adoption of 4G specific agricultural technology. Our proxy for 4G adoption is installation of agricultural applications that improve information access among farmers. Specifically, we exploit a proprietary dataset of geocoded and time-stamped mobile application ("app") installations from Krishify – an Indian firm which runs one of the largest social networks and commerce platforms for farmers. Appendix section D.1 presents a detailed description of app and the dataset.

Our dataset of Krishify app installation contains geolocations of users along with the dates on which they installed the app. The earliest date of app installation is May 2019. Thus, we subset our hexagon-level dataset to keep hexagons which got treated after May 2019 i.e. received their first 4G BTS after May 2019. We superimpose the geolocations of users' app installations on our hexagons and calculate the number of monthly installations at the hexagon level.

Using dates on which the first 4G BTS became operational as the treatment date, we run a difference-in-differences regression at the hexagonal level to study the effect of 4G *availability* on Krishify app installations. Figure 16 presents the results using the Callaway and Sant'Anna (2021) estimator. We observe positive, statistically significant, and persistent growth in the number of installations starting from the first month after treatment. Moreover, we observe no significant pre-trends in the months leading up to the treatment. On the extensive margin we find that app installations increase by 20% two years after the introduction of 4G (see Figure 16a). The total number of app installations increases to 4 installations per month at the hexagon level (Figure 16b)). The increase in the number of installations is equivalent to 15

correlation coefficient of 0.632. We think of these two measures as alternative ways to target a similar economic parameter, i.e., the trust in institutions that are the primary modes of the traditional information dissemination processes.

times the pre-treatment period monthly average (Figure 16c).[26] Lastly, we show in appendix Figure D.4 that we do not observe a positive significant effect on Krishify app installations following the introduction of 3G towers. This result makes explicit what has been implicit in our analysis till now – availability of 4G is associated with 4G adoption, as proxied through the installation of Krishify app.

7.1 Technology Adoption: What type of information do farmers access?

This section examines the type of information farmers access on the internet using detailed browsing and activity data of farmers on the Krishify app. The objective of this section is to present direct evidence suggesting that internet-based app allows farmers to access relevant information.

We use a 10% random sample of all farmers in Krishify. This random sample includes over 400,000 farmers and provides details of the browsing, search, likes, and comment history of all farmers since they joined the application. Additionally, we can also observe information on their land holdings and their risk-taking tendency. Appendix section D.1 presents a detailed description of the dataset.

We classify all interactions of these farmers into 13 distinct topics which include – agricultural news, information about government programs, farm machinery, crop management, weather, animal husbandry, fertilizer, seeds, prices, and other inputs, as well as content that is recreational and political. Table 3 presents the list of these topics.

The second column of Table 3 presents the percentage of interactions for each topic among our sample farmers. The foremost four topics collectively constitute about 70% of farmer interactions on the platform. Notably, government programs rank highest, contributing to 21.3% of all interactions, followed by agricultural news at 16.6%, information about farm machinery (e.g., tractors) at 16.1%, and weather information at 14.2%. Furthermore, farmers also utilize the platform to seek advice specific to animal (9.1%) and crop (8.18%) management. Interactions related to inputs like fertilizers, seeds, finance, and other miscellaneous inputs collectively make up 11.4% of all farmer engagements on the platform.

Columns (2)-(6) and columns (7)-(8) report numbers for farmers with different land ownership and risk-taking abilities, respectively, and find that the topic-wise interaction of farmers is similar across these divisions.

[26] We find similar results using log(1+# Installations) and inverse hyperbolic sine transformation of the number of installations. See Appendix Figure D.3.

The detailed activity data of farmers suggests that farmers may be using internet-based platforms extensively for gathering information on a wide variety of agri-related activities. This descriptive analysis, coupled with the findings presented in section 7, reinforces the confidence in our main finding that the introduction of 4G technology enhanced information accessibility for farmers.

7.2 What type of farmers are more likely to use such services?

The objective of this section is to discuss some correlations indicating what type of farmers are more likely to use high-speed internet-powered services. Specifically, we examine the correlation between the activity on the Krishify app and farmer characteristics such as land-holding and risk-taking ability. We use three different measures of the level of adoption – (1) engagement score which measures the total activity of the farmer on the app, (2) direct measures of active engagement such as likes, comments, and videos posted, and (3) timing of adoption.

First, we use the engagement score to measure the amount of total activity of the farmer on the app. Appendix Figure D.8 presents the kernel density of the engagement score. Table 4 presents the results from the Poisson regression using the engagement score as the dependent variable.[27] Two key facts emerge. First, land-owning farmers have greater total engagement relative to landless or tenant farmers. However, conditional on owning land, small and medium farmers have a greater engagement relative to large farmers with land above ten acres. Second, farmers with higher risk-taking ability have greater overall engagement.

Next, we employ the number of likes, comments and videos posted on the application as a measure of active engagement. Table 5 presents the results from the Poisson regression using these measures of active engagement on the platform as the dependent variable. All specifications include cohort × district fixed effects effectively comparing farmers in the same district that installed the application during the same month. Farmers with higher levels of risk-taking ability are associated with higher levels of active engagement. Farmers who own less than two acres of land show lower levels of active engagement compared to landless or tenant farmers. However, active engagement tends to increase in hold-holding as land holding size surpasses two acres. This relationship between land-holding and active engagement is different from the relationship between land-holding and total engagement presented in Table

[27]Note that we include cohort fixed effects. This allows us to compare engagement scores among farmers who installed Krishify in the same month. Including this fixed effect is crucial because the engagement score might be mechanically higher for farmers who have been using the platform for a longer duration. However, for completeness, we present results both with and without cohort-fixed effects.

4, indicating that land-holding size may be associated with the type of engagement on internet-based platforms.

Lastly, we use the time to adoption of the application as a measure of adoption. Time to adoption is measured as the duration between the installation date of the app by a farmer and the date of the first app installation in the same district. Appendix Figure D.9 presents the kernel density of the measure. This measure indicates technology adoption on an extensive margin, relative to the peers. Table 6 presents the results from the Poisson regression using the time to adoption as the dependent variable.[28] Results indicate that landless or tenant farmers are faster to adopt relative to land-owning farmers. Moreover, farmers with greater risk-taking ability take longer to adopt as well.

8 Conclusion

Can high-speed internet boost information access and enhance productivity? We examine this question in the context of Indian agriculture where information frictions, especially frictions related to information dissemination, are rampant. Combining granular geographic data on the introduction of 4G with remote-sensing data on agricultural productivity we show that the improvement in information dissemination due to the introduction of 4G leads to improvement in productivity.

Our analysis reveals that the introduction of high-speed 4G internet is associated with significant improvements in agricultural outcomes, including - (1) a 1.06% increase in the yields of cereal crops, (2) a 4.28% increase in the production value of cereal crops, and (3) a 3.21% increase in the production value of all crops. Six years post the introduction of 4G internet, agricultural households experienced a substantial 14.50% growth in annual income. Our calculations attribute 24.56% of this income growth from 2014 to 2021 to the improved accessibility of information. Moreover, we find that 4G introduction leads to increased fertilizer consumption, convergence towards optimal N:P:K ratios, and greater credit uptake.

Our study indicates that the effectiveness of information dissemination is contingent on the usefulness, reliability, accuracy, and value of the information. Additionally, areas with lower trust in local and state institutions, experience a stronger impact from the introduction of high-speed 4G internet.

Overall our results indicate that while high-speed internet acts as an important tool for

[28]Note that we are unable to include cohort fixed effects in this analysis as by construction time to adoption is highly collinear with cohort fixed effects and exhibits little heterogeneity within a cohort.

information dissemination, our findings underscore the importance of considering the quality of information. Merely introducing internet infrastructure may not be sufficient; instead, it should be viewed as a complement to other factors. Thus, comprehensive infrastructure development that enhances information accessibility along with the availability, reliability, and accuracy of information holds the key to maximizing the effect.

References

Aker, Jenny C, and Isaac M Mbiti. 2010. "Mobile phones and economic development in Africa." *Journal of economic Perspectives*, 24(3): 207–232.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. "The skill complementarity of broadband internet." *The Quarterly Journal of Economics*, 130(4): 1781–1824.

Andersen, Thomas Barnebeck, Jeanet Bentzen, Carl-Johan Dalgaard, and Pablo Selaya. 2011. "Does the Internet reduce corruption? Evidence from US states and across countries." *The World Bank Economic Review*, 25(3): 387–417.

Anderson, Jock R, and Gershon Feder. 2004. "Agricultural extension: Good intentions and hard realities." *The World Bank Research Observer*, 19(1): 41–60.

Asher, Sam, and Paul Novosad. 2020. "Rural roads and local economic development." *American economic review*, 110(3): 797–823.

Batista, Catia, and Pedro C Vicente. 2023. "Is mobile money changing rural Africa? Evidence from a field experiment." *Review of Economics and Statistics*, 1–29.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2013. "Profitability of fertilizer: Experimental evidence from female rice farmers in Mali." *American Economic Review*, 103(3): 381–386.

Birch, Colin PD, Sander P Oom, and Jonathan A Beecham. 2007. "Rectangular and hexagonal grids used for observation, experiment and simulation in ecology." *Ecological modelling*, 206(3-4): 347–359.

Bloom, Nicholas. 2009. "The impact of uncertainty shocks." *econometrica*, 77(3): 623–685.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2022. "Revisiting Event Study Designs: Robust and Efficient Estimation."

Braghieri, Luca, Roee Levy, and Alexey Makarin. 2022. "Social media and mental health." *American Economic Review*, 112(11): 3660–3693.

Brooks, Wyatt, and Kevin Donovan. 2020. "Eliminating uncertainty in market access: The impact of new bridges in rural Nicaragua." *Econometrica*, 88(5): 1965–1997.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Camacho, Adriana, and Emily Conover. 2010. "The impact of receiving price and climate information in the agricultural sector." *Documento CEDE*, , (2010-40).

Casaburi, Lorenzo, Michael Kremer, Sendhil Mullainathan, and Ravindra Ramrattan. 2014. "Harnessing ICT to increase agricultural production: Evidence from Kenya." *Harvard University*.

Cole, Shawn, and A Nilesh Fernando. 2021. "Mobileizing agricultural advice technology adoption diffusion and sustainability." *The Economic Journal*, 131(633): 192–219.

Cole, Shawn, and G Sharma. 2017. "The promise and challenges in implementing ICT for agriculture." Working paper. Available at: https://precisionag. org/uploads/cole-sharma .

Cole, Shawn, and Wentao Xiong. 2017. "Agricultural insurance and economic development." *Annual review of Economics*, 9: 235–262.

Cole, Shawn, Xavier Giné, and James Vickery. 2017. "How does risk management influence production decisions? Evidence from a field experiment." *The Review of Financial Studies*, 30(6): 1935–1970.

Deaton, Angus. 1990. "On Risk, Insurance and Intra-Village Consumption Smoothing." *Research Program in Development Studies. Princeton University*.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–2996.

de Chaisemartin, Clément, and Xavier D'Haultfuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya." *American economic review*, 101(6): 2350–2390.

DAndrea, Angelo, and Nicola Limodio. 2023. "High-Speed Internet, Financial Technology, and Banking." *Management Science*.

Emerick, Kyle, Alain De Janvry, Elisabeth Sadoulet, and Manzoor H Dar. 2016. "Technological innovations, downside risk, and the modernization of agriculture." *American Economic Review*, 106(6): 1537–1561.

Fabregas, Raissa, Michael Kremer, and Frank Schilbach. 2019. "Realizing the potential of digital development: The case of agricultural advice." *Science*, 366(6471): eaay3038.

Fafchamps, Marcel, and Bart Minten. 2012. "Impact of SMS-based agricultural information on Indian farmers." *The World Bank Economic Review*, 26(3): 383–414.

Ghosh, Pulak, and Nishant Vats. 2022. "Safety Nets, Credit, and Investment: Evidence from a Guaranteed Income Program." *University of Chicago Working Paper*.

Gonzalez, Robert M. 2021. "Cell phone access and election fraud: evidence from a spatial regression discontinuity design in Afghanistan." *American Economic Journal: Applied Economics*, 13(2): 1–51.

GSMA. 2020. "Rights of Way (RoW) in India: A submission from the GSMA relating to India Mobile Sector." GSMA.

Guiteras, Raymond. 2009. "The impact of climate change on Indian agriculture." *Working Paper, University of Maryland*.

Gupta, Apoorv, Jacopo Ponticelli, and Andrea Tesei. 2023*a*. "Information Frictions and Take-up of Government Credit Programs."

Gupta, Apoorv, Jacopo Ponticelli, and Andrea Tesei. 2023*b*. "Language Barriers, Technology Adoption and Productivity: Evidence from Agriculture in India."

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. 2021. "3g internet and confidence in government." *The Quarterly Journal of Economics*, 136(4): 2533–2613.

Hill, Ruth Vargas, Miguel Robles, and Francisco Ceballos. 2016. "Demand for a simple weather insurance product in India: theory and evidence." *American Journal of Agricultural Economics*, 98(4): 1250–1270.

Hjort, Jonas, and Jonas Poulsen. 2019. "The arrival of fast internet and employment in Africa." *American Economic Review*, 109(3): 1032–1079.

Huete, Alfredo, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao, and Laerte G Ferreira. 2002. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote sensing of environment*, 83(1-2): 195–213.

Jayachandran, Seema. 2006. "Selling labor low: Wage responses to productivity shocks in developing countries." *Journal of political Economy*, 114(3): 538–575.

Kaur, Supreet. 2019. "Nominal wage rigidity in village labor markets." *American Economic Review*, 109(10): 3585–3616.

Manacorda, Marco, and Andrea Tesei. 2020. "Liberation technology: Mobile phones and political mobilization in Africa." *Econometrica*, 88(2): 533–567.

Mehrabi, Zia, Mollie J McDowell, Vincent Ricciardi, Christian Levers, Juan Diego Martinez, Natascha Mehrabi, Hannah Wittman, Navin Ramankutty, and Andy Jarvis. 2021. "The global divide in data-driven farming." *Nature Sustainability*, 4(2): 154–160.

Mobarak, Ahmed Mushfiq, and Mark R Rosenzweig. 2013. "Informal risk sharing, index insurance, and risk taking in developing countries." *American Economic Review*, 103(3): 375–380.

Robles, Miguel, et al. 2021. "Agricultural insurance for development: Past, present, and future." *Agricultural development: New perspectives in a changing world*, 563–94.

Rosenzweig, Mark, and Christopher R Udry. 2013. "Forecasting profitability." National Bureau of Economic Research.

Roth, Jonathan. 2024. "Interpreting Event-Studies from Recent Difference-in-Differences Methods." *arXiv preprint arXiv:2401.12309*.

Roth, Jonathan, Pedro HC SantAnna, Alyssa Bilinski, and John Poe. 2023. "Whats trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics*.

Stiglitz, Joseph E. 2000. "The contributions of the economics of information to twentieth century economics." *The quarterly journal of economics*, 115(4): 1441–1478.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Suri, Tavneet. 2017. "Mobile money." *Annual Review of Economics*, 9: 497–520.

Suri, Tavneet, and William Jack. 2016. "The long-run poverty and gender impacts of mobile money." *Science*, 354(6317): 1288–1292.

Townsend, Robert M. 1994. "Risk and insurance in village India." *Econometrica: journal of the Econometric Society*, 539–591.

Udry, Christopher, Federica di Battista, Mathias Fosu, Markus Goldstein, Alev Gurbuz, Dean Karlan, and Shashidhara Kolavalli. 2019. "Information, market access and risk: Addressing constraints to agricultural transformation in northern Ghana." *Draft Report*.

Van Campenhout, Bjorn, David J Spielman, and Els Lecoutere. 2021. "Information and communication technologies to provide agricultural advice to smallholder farmers: Experimental evidence from Uganda." *American Journal of Agricultural Economics*, 103(1): 317–337.

Vasilaky, Kathryn, Kentaro Toyama, Tushi Baul, Mohak Mangal, and Urmi Bhattacharya. 2015. "Learning Digitally: Evaluating the Impact of Farmer Training via Mediated Videos."

Zuo, George W. 2021. "Wired and hired: Employment effects of subsidized broadband Internet for low-income Americans." *American Economic Journal: Economic Policy*, 13(3): 447–482.

Table 1: Summary Statistics

Panel A: Agricultural Productivity Measures	# Obs	Min	p25	p50	p75	Max	Mean	SD
EVI Implied Yield	255,330	0.000	0.165	0.267	0.369	0.700	0.270	0.144
Maximum Yield	255,330	0.158	0.511	0.602	0.676	0.896	0.584	0.138

Panel B: Unit (Hexagon) Level Measures	# Obs	Min	p25	p50	p75	Max	Mean	SD
Agricultural Landcover	28,370	1.000	4.000	4.000	7.000	8.000	5.064	2.251
Dist. to Weather Station (km)	28,370	4.863	25.770	38.916	52.116	92.391	39.830	18.864
Coef of Var (Monsoon Rainfall)	22,251	0.127	0.269	0.354	0.520	1.414	0.427	0.246
Prob(Predictable Weather)	23,079	0.000	0.000	0.038	0.197	0.604	0.107	0.142
Prob(Difficult-to-predict Weather)	23,079	0.000	0.000	0.000	0.006	0.483	0.023	0.072

Panel C: District Level Measures	# Obs	Min	p25	p50	p75	Max	Mean	SD
Trust (Village Panchayat)	382	0.056	0.368	0.447	0.553	0.967	0.463	0.141
Trust (State Government)	382	0.070	0.374	0.451	0.567	0.906	0.468	0.148
LN(Gross Sown Area, GSA in hectares)	5,649	1.723	4.873	5.301	5.808	7.400	5.277	0.780
Total Fertilizer (in ’000 kg)/GSA	5,649	0.189	63.227	124.625	221.007	789.677	155.147	121.657
Nitrogen Fertilizer (in ’000 kg)/GSA	5,649	0.126	37.432	75.876	144.535	515.728	101.102	85.767
Phosphate Fertilizer (in ’000 kg)/GSA	5,649	0.031	17.523	32.564	55.007	232.372	39.230	29.664
Potash Fertilizer (in ’000 kg)/GSA	5,649	0.005	2.126	7.440	17.732	139.473	14.230	18.905
Distance to Optimal NPK Ratio	5,649	0.011	0.108	0.165	0.218	0.454	0.168	0.080

Panel D: Branch-Level Agricultural Credit	# Obs	Min	p25	p50	p75	Max	Mean	SD
LN(Agricultural Credit)	133,056	12.2326	16.9629	18.0691	19.0249	20.5999	17.8341	1.5743

This table reports the descriptive statistics for the key variables used in the analysis. Panel A reports the summary statistics for agricultural productivity variables, and panel B reports the summary statistics for unit (hexagon) level variables. Panel C reports variables measured at the district level. Panel D reports the summary statistics for branch-level agricultural credit. All variables are winsorized at the 1% level.

Table 2: Aggregate Average Treatment Effect on the Treated (ATT)

Specification	Partially Aggregated							ATT
Simple								0.053**
								(0.022)
Group-specific	g=2014	g=2015	g=2016	g=2017	g=2018	g=2019	g=2020	
	0.050	0.053*	0.082**	0.078**	-0.123**	0.057	-0.105	0.047**
	(0.035)	(0.032)	(0.039)	(0.039)	(0.054)	(0.055)	(0.075)	(0.022)
Event study	e = 0	e = 1	e = 2	e = 3	e = 4	e = 5	e = 6	
	0.034**	-0.033*	0.014	0.090**	0.145***	0.186***	0.245***	0.097***
	(0.015)	(0.019)	(0.028)	(0.037)	(0.048)	(0.057)	(0.073)	(0.028)

This table reports the aggregate treatment effect parameters with standard errors clustered at the district level shown in parenthesis. The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. The dependent variable is standardized to mean zero and standard deviation of one. Row "Simple" reports the weighted average of all group-time $ATT(g,t)$'s as defined in Callaway and Sant'Anna (2021). Rows "Group-specific" and "Event study" list the partially aggregated group-specific and event-time specific average treatment effects where g indexes the year of treatment of the group and e indexes event-time. The last column of each row summarizes the aggregate average treatment effect on the treated (ATT) by taking a weighted average of partially aggregated group-specific and event-time specific average treatment effects following Callaway and Sant'Anna (2021). *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Topic-wise Interaction of Farmers on Krishify

Topic	Overall	By Land Ownership					By Risk-Taking Ability		
		No Land	Below 2 Acres	2-5 Acres	5-10 Acres	Above 10 Acres	Low	Medium	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Government Schemes	21.29%	21.24%	21.98%	20.70%	21.13%	20.52%	21.46%	22.71%	19.60%
Agriculture News	16.65%	15.27%	18.02%	17.76%	18.28%	17.24%	17.64%	15.34%	15.79%
Farm Machinery	16.09%	15.92%	15.05%	16.62%	18.09%	17.85%	18.12%	13.73%	14.03%
Weather Information	14.24%	16.68%	11.76%	11.79%	11.56%	14.35%	16.86%	10.95%	11.82%
Animal Husbandry	9.07%	8.39%	11.78%	8.55%	7.38%	6.84%	5.30%	14.10%	12.27%
Crop Management	8.18%	7.41%	8.23%	9.74%	9.08%	9.16%	7.01%	8.69%	10.15%
Fertilizer	5.15%	5.34%	4.44%	5.69%	5.20%	5.15%	4.96%	4.86%	5.81%
Finance	2.96%	2.67%	3.43%	3.09%	3.08%	2.90%	2.20%	3.99%	3.60%
Seeds	2.02%	2.20%	1.69%	2.06%	1.91%	1.94%	1.98%	1.73%	2.37%
Price Information	1.39%	1.53%	1.08%	1.28%	1.59%	1.46%	1.45%	1.27%	1.36%
Other Inputs	1.31%	1.46%	1.08%	1.27%	1.22%	1.15%	1.27%	1.14%	1.53%
Recreational	0.89%	0.92%	0.84%	0.91%	0.92%	0.73%	0.82%	0.95%	0.97%
Politics	0.77%	0.96%	0.63%	0.54%	0.57%	0.70%	0.91%	0.54%	0.71%

This table presents the percentage of interactions for each topic among our sample farmers. We classify all interactions of these farmers into 13 distinct topics which include – agricultural news, information about government programs, farm machinery, crop management, weather, animal husbandry, fertilizer, seeds, prices, and other inputs, as well as content that is recreational and political. Column (1) reports the numbers of all farmers. Columns (2)-(6) present the numbers by different land-ownership buckets. Land ownership is divided into five buckets no land ownership, land below 2 acres, land between 2 to 5 acres, land between 5 to 10 acres and land above 10 acres. Columns (7)-(9) reports the numbers for farmers in different buckets of risk-taking ability. Risk-taking ability is divided into three categories of low, medium, and high.

Table 4: Level of Total Engagement & Farmer Characteristics

Dep Var: Engagement Score	(1)	(2)	(3)	(4)	(5)	(6)
Land, Below 2 Acres	0.2043***		0.1558***	0.0696***	0.0638***	0.0631***
	(0.0037)		(0.0036)	(0.0037)	(0.0037)	(0.0037)
Land, 2-5 Acres	0.1986***		0.1681***	0.0821***	0.0779***	0.0776***
	(0.0047)		(0.0047)	(0.0049)	(0.0048)	(0.0048)
Land, 5-10 Acres	0.1711***		0.1524***	0.0658***	0.0678***	0.0677***
	(0.0056)		(0.0056)	(0.0055)	(0.0054)	(0.0055)
Land, Above 10 Acres	0.1033***		0.0926***	0.006	0.0172***	0.0175***
	(0.0062)		(0.0059)	(0.0061)	(0.0059)	(0.0060)
Risk-Taking Ability, Medium		0.2516***	0.2250***	0.2232***	0.2173***	0.2160***
		(0.0041)	(0.0040)	(0.0040)	(0.0039)	(0.0038)
Risk-Taking Ability, High		0.3615***	0.3277***	0.3233***	0.3108***	0.3100***
		(0.0094)	(0.0092)	(0.0092)	(0.0087)	(0.0087)
Cohort FE				Yes	Yes	
District FE					Yes	
District X Cohort FE						Yes
# Obs	405,693	405,693	405,693	405,693	405,693	405,693

This table presents the relationship between farmer-level engagement on the Krishify application and their characteristics. The dependent variable is the engagement score that is provided by Krishify and measures the total amount of engagement of the farmer on the platform. Land ownership is divided into five buckets no land ownership, land below 2 acres, land between 2 to 5 acres, land between 5 to 10 acres and land above 10 acres. For land ownership, farmers with no land ownership are the omitted category. Risk-taking ability is divided into three categories of low, medium, and high. Low risk-taking ability is the omitted category. We use Poisson regressions to estimate the relationship. Cohort refers to the cohort of farmers that install the app during the same month. District refers to the district of farmer location. All variables are winsorized at the 1% level. Standard errors in parentheses are estimated by clustering at the district level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Level of Active Engagement & Farmer Characteristics

	(1)	(2)	(3)
	# Likes	# Comments	# Videos Posted
Land, Below 2 Acres	-0.0415***	-0.0373***	-0.0396
	(0.0130)	(0.0140)	(0.0429)
Land, 2-5 Acres	0.0905***	0.1006***	0.0795*
	(0.0152)	(0.0147)	(0.0478)
Land, 5-10 Acres	0.1159***	0.1225***	0.2170***
	(0.0173)	(0.0181)	(0.0578)
Land, Above 10 Acres	0.1466***	0.0474**	0.2648***
	(0.0188)	(0.0187)	(0.0587)
Risk-Taking Ability, Medium	0.7682***	1.1795***	1.3651***
	(0.0137)	(0.0131)	(0.0377)
Risk-Taking Ability, High	0.9317***	1.2855***	1.4531***
	(0.0142)	(0.0148)	(0.0351)
District X Cohort FE	Yes	Yes	Yes
# Obs	402,828	396,728	276,631
Sample Dep Var Mean	1.6178	0.3676	0.0193
Sample Dep Var SD	4.5000	1.0400	0.1377

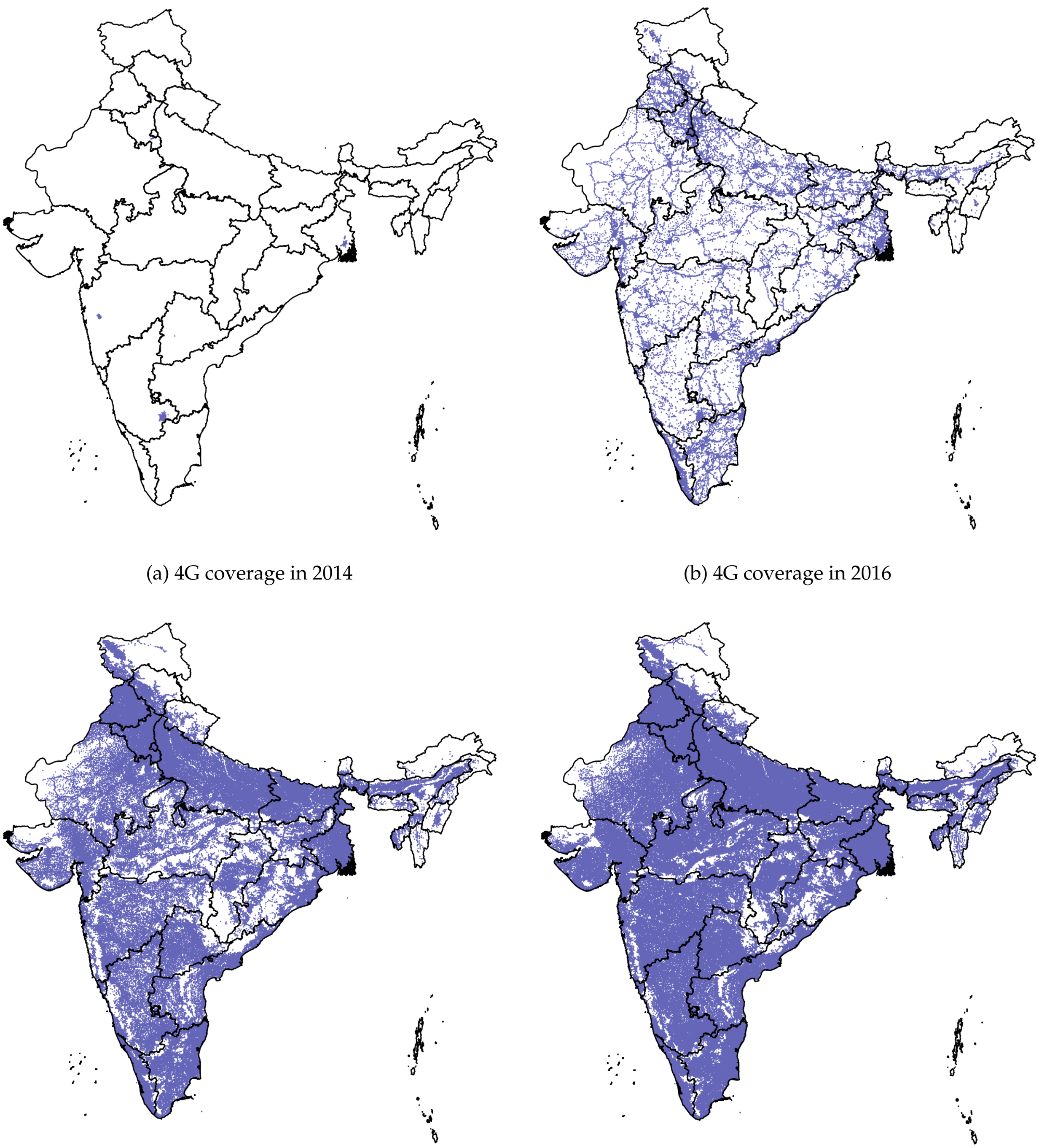
This table presents the relationship between farmer-level active engagement on the Krishify application and their characteristics. The dependent variable is the total number of likes in column (1), total number of comments in column (2), and total number of videos posted in column (3). Land ownership is divided into five buckets no land ownership, land below 2 acres, land between 2 to 5 acres, land between 5 to 10 acres and land above 10 acres. For land ownership, farmers with no land ownership are the omitted category. Risk-taking ability is divided into three categories of low, medium, and high. Low risk-taking ability is the omitted category. We use Poisson regressions to estimate the relationship. Cohort refers to the cohort of farmers that install the app during the same month. District refers to the district of farmer location. All variables are winsorized at the 1% level. Standard errors in parentheses are estimated by clustering at the district level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Time to Adoption & Farmer Characteristics

Dep Var: Time to Adoption (in days)	(1)	(2)	(3)	(4)
Land, Below 2 Acres	0.6358***		0.6315***	0.6305***
	(0.0033)		(0.0032)	(0.0030)
Land, 2-5 Acres	0.6295***		0.6269***	0.6233***
	(0.0034)		(0.0034)	(0.0033)
Land, 5-10 Acres	0.6331***		0.6315***	0.6265***
	(0.0036)		(0.0035)	(0.0039)
Land, Above 10 Acres	0.6335***		0.6326***	0.6234***
	(0.0040)		(0.0040)	(0.0042)
Risk-Taking Ability, Medium		0.1165***	0.0218***	0.0219***
		(0.0043)	(0.0027)	(0.0026)
Risk-Taking Ability, High		0.1319***	0.0273***	0.0308***
		(0.0043)	(0.0029)	(0.0027)
District FE				Yes
# Obs	405,717	405,717	405,717	405,717

This table presents the relationship between farmer-level time to adoption of the Krishify application and their characteristics. The dependent variable, time to adoption, is measured as the duration between the installation date of the app by a farmer and the date of the first app installation in the same district. Land ownership is divided into five buckets no land ownership, land below 2 acres, land between 2 to 5 acres, land between 5 to 10 acres and land above 10 acres. For land ownership, farmers with no land ownership are the omitted category. Risk-taking ability is divided into three categories of low, medium, and high. Low risk-taking ability is the omitted category. We use Poisson regressions to estimate the relationship. District refers to the district of farmer location. All variables are winsorized at the 1% level. Standard errors in parentheses are estimated by clustering at the district level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Figure 1: Evolution of 4G coverage over time



(a) 4G coverage in 2014

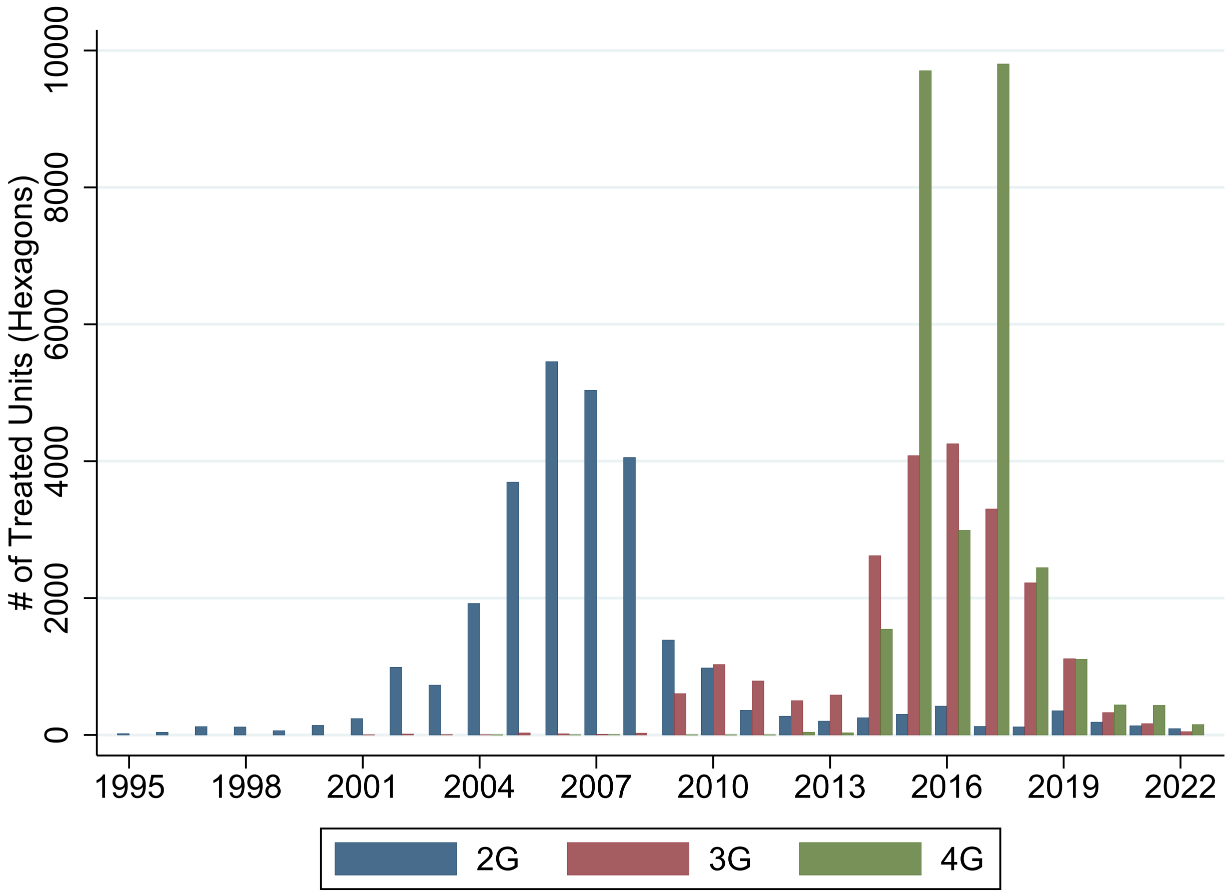
(b) 4G coverage in 2016

(c) 4G coverage in 2018

(d) 4G coverage in 2020

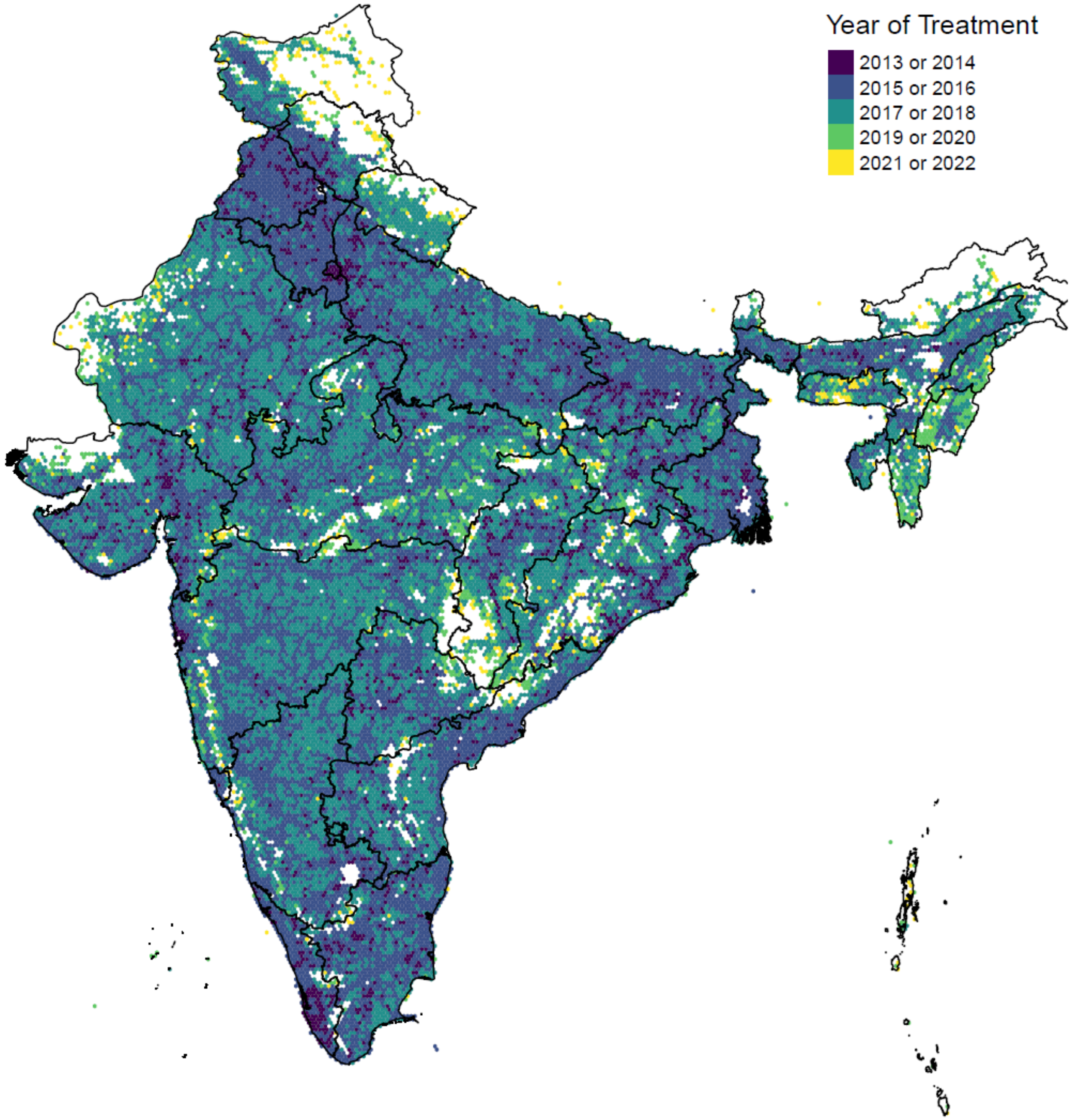
This figure plots the evolution of 4G coverage over time based on the Ministry of Telecom data.

Figure 2: Staggered introduction of BTS by Technology



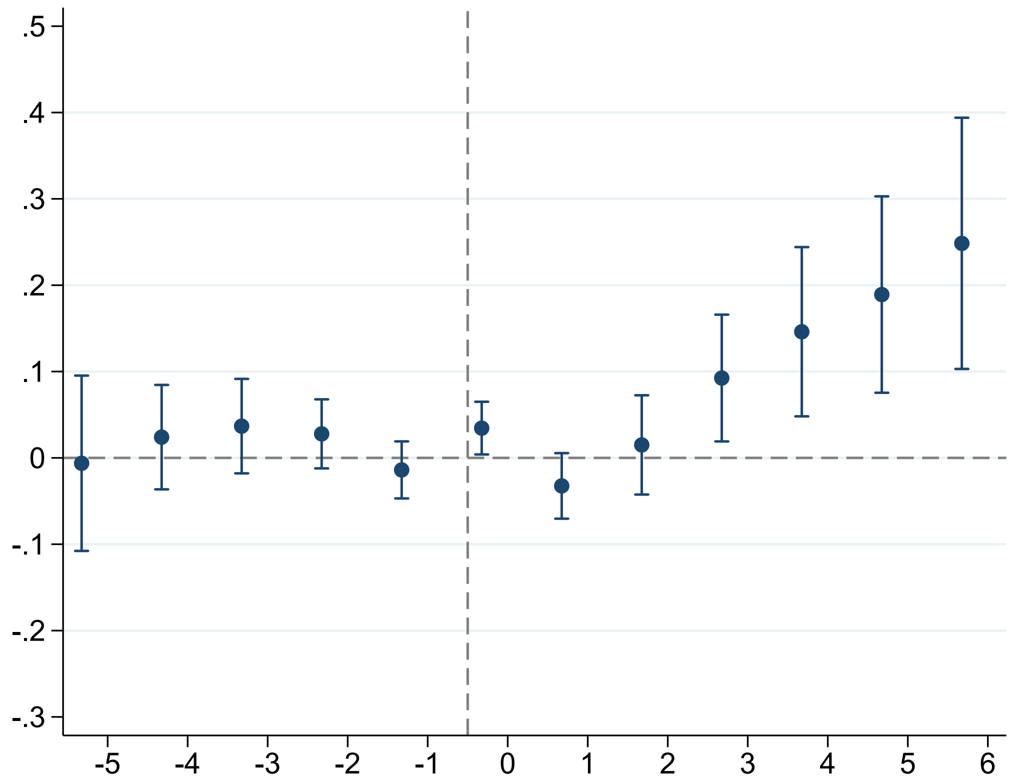
This figure presents the staggered introduction of BTs by technology type from 1995 until 2022. The X-axis denotes the calendar year. The Y-axis denotes the number of units (hexagons) treated by technology type for each year. The three technology types include 2G, 3G, and 4G.

Figure 3: Evolution of 4G Treatment Across Hexagons



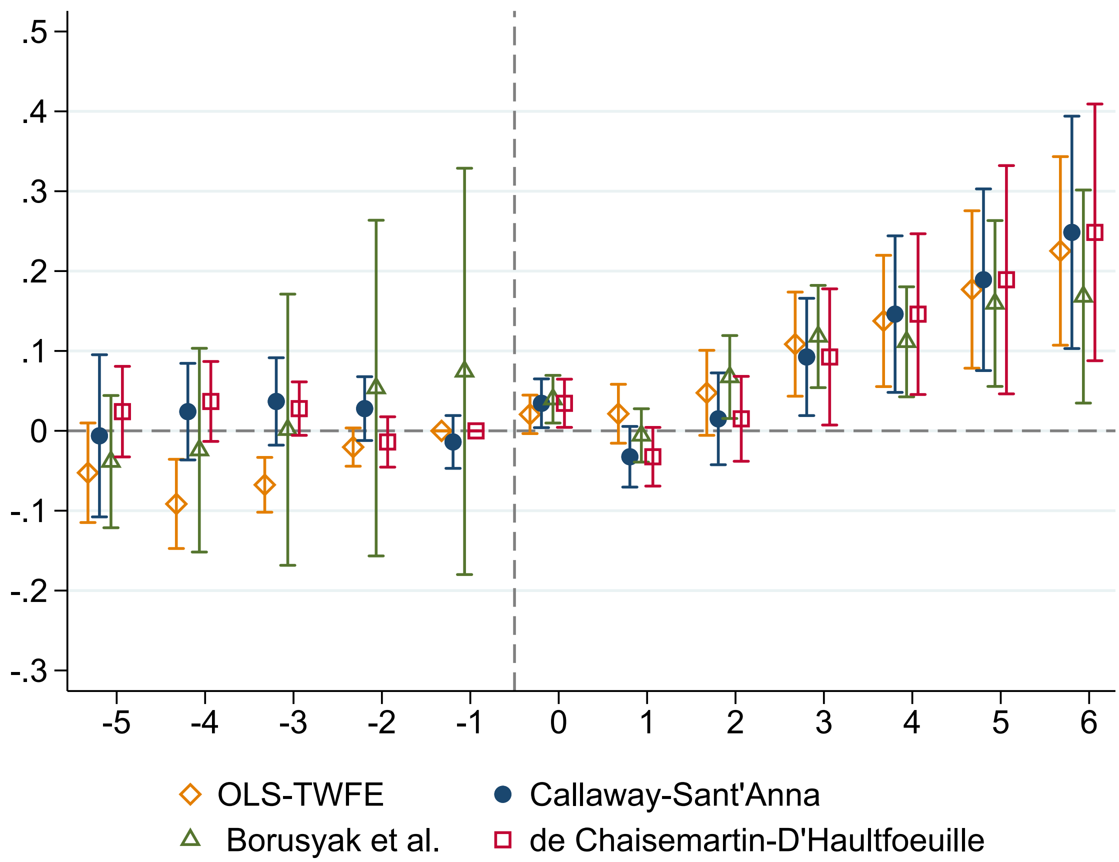
This figure presented a tessellated heat map of India. Each color denotes the year in which the 4G technology was introduced in the hexagon for the first time.

Figure 4: Effect of 4G introduction on EVI implied agricultural yield



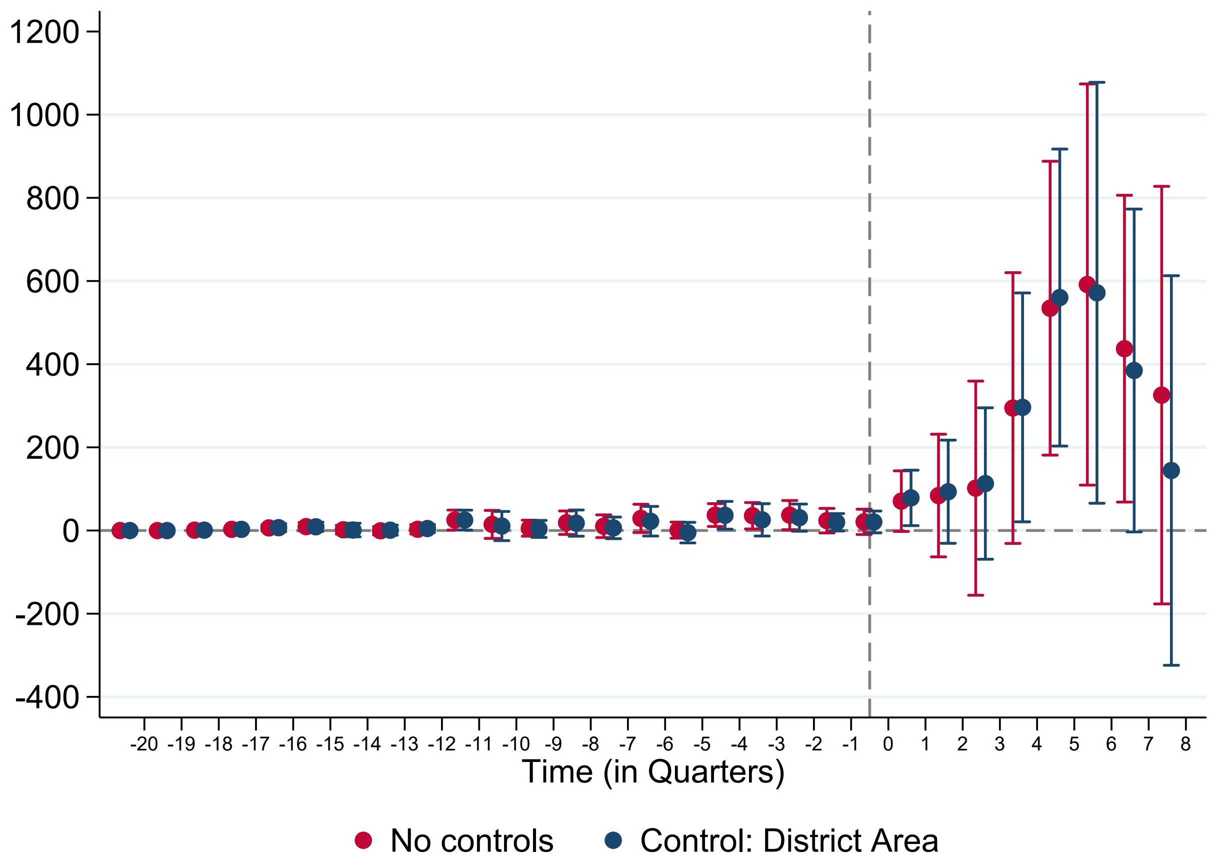
DiD estimates for our baseline measure – EVI implied agricultural yield: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. The dependent variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 5: Robustness using alternative DiD estimators: Baseline Measure



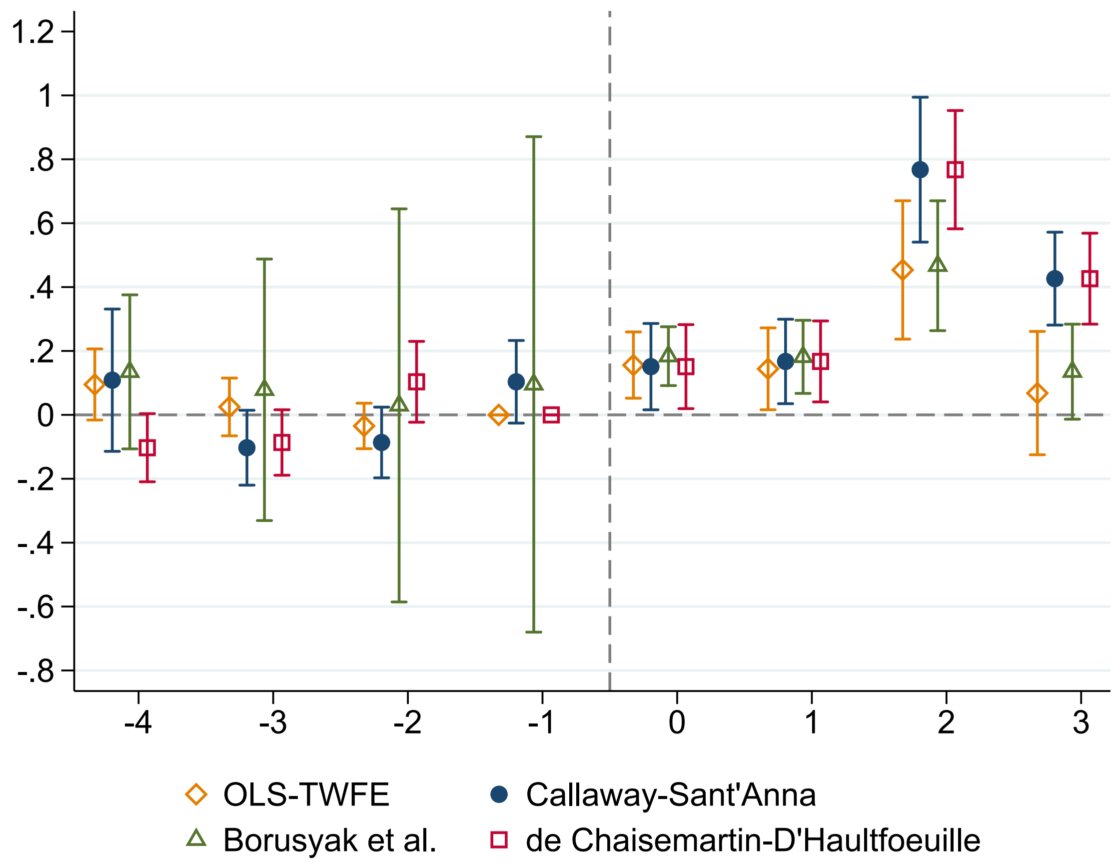
Robustness with multiple DiD estimators for our baseline measure – EVI implied agricultural yield This figure plots the event-study dynamic coefficients using four different estimators: (1) dynamic TWFE estimator (in orange with hollow diamond markers), (2) Callaway and Sant'Anna (2021) estimator (in blue with solid circle markers), (3) Borusyak, Jaravel and Spiess (2022) estimator (in green with hollow triangle markers), and (4) De Chaisemartin and d'Haultfoeuille (2020) estimator (in pink with hollow square markers). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 6: Effect of RoW Adoption on Installation of 4G Telecom Towers



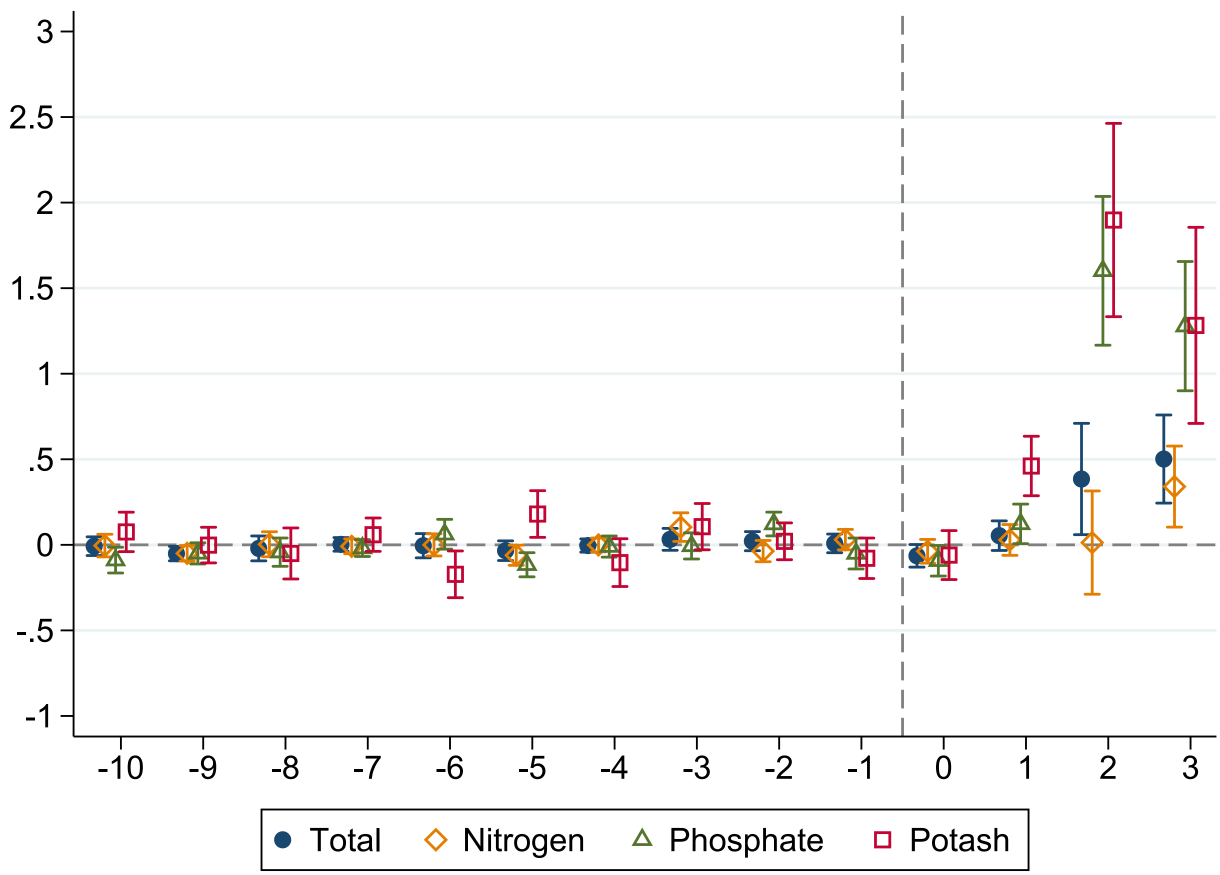
DiD estimates for the impact of staggered adoption of Rights of Way (RoW) rules across states on number of BTSs (first-stage results for RoW-policy based identification strategy): This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is the quarterly stock of the number of 4G BTSs operational at the district-level. Consequently, event-time is measured in quarters. Treatment timing is constructed using the dates of introduction of RoW rules across different States/Union Territories as listed in C.2. The sample consists of the districts that fall within States/Union Territories which passed RoW rules from 2015 to 2019. All measures are winsorized at 1% level. Standard errors are clustered at the state level. The bars represent 95 percent confidence intervals.

Figure 7: Effect of RoW Adoption on Agricultural Production



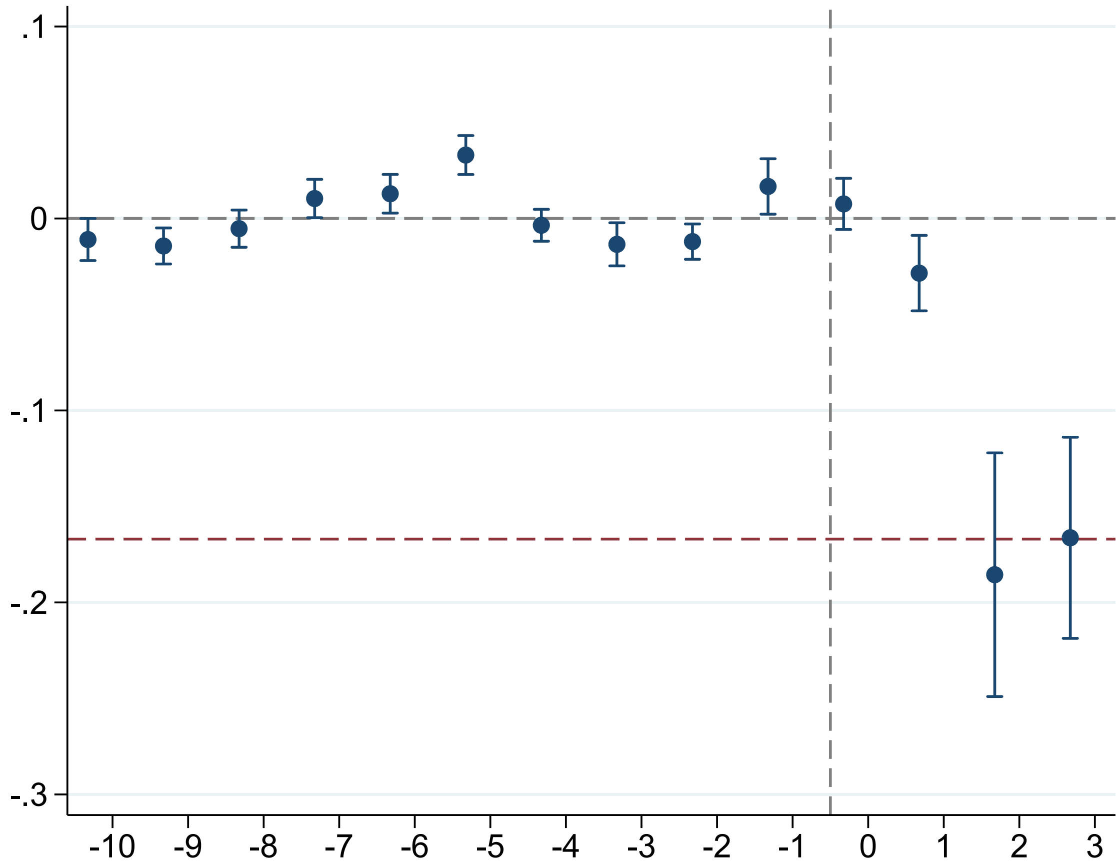
DiD estimates for the impact of (staggered) passage of Rights of Way (RoW) rules across States on EVI implied agricultural yields (second-stage results for RoW-policy based identification strategy): This figure plots the event-study coefficients estimated using four different estimators: (1) dynamic TWFE estimator (in orange with hollow diamond markers), (2) Callaway and Sant'Anna (2021) estimator (in blue with solid circle markers), (3) Borusyak, Jaravel and Spiess (2022) estimator (in green with hollow triangle markers), and (4) De Chaisemartin and d'Haultfoeuille (2020) estimator (in pink with hollow square markers). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. Treatment timing is constructed using the dates of introduction of RoW rules across different States/Union Territories as listed in C.2. The sample consists of the subset of hexagons that fall within States/Union Territories which passed RoW rules from 2015 to 2019, and saw the introduction of 4G BTS between 2013 and 2021. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 8: 4G introduction & Fertilizer Usage



DiD estimates for impact of (staggered) passage of Rights of Way (R0W) rules across states on fertilizer consumption: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variables are measured as the natural logarithm of the amount of consumption of total fertilizers (NPK, in blue with solid circle markers), nitrogen (N, in orange with hollow diamond markers), phosphate (P, in green with hollow triangle markers), and potash (K, in pink with hollow square markers) per unit of gross sown area (GSA). All variables are measured at the district-year level from 1998 until 2020. Treatment timing is constructed using the dates of introduction of RoW rules across different States/Union Territories as listed in C.2. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 9: 4G introduction & Optimal NPK Ratio

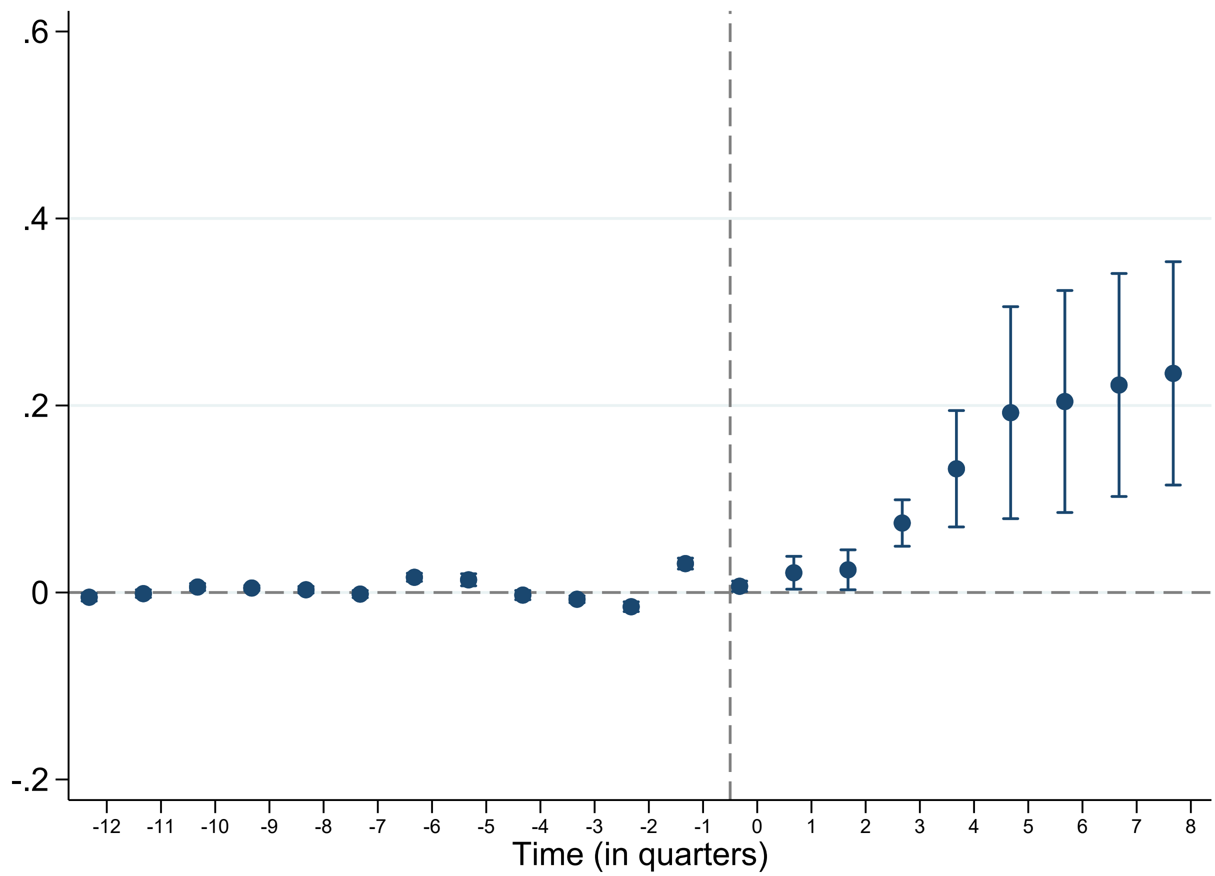


DiD estimates for impact of (staggered) passage of Rights of Way (R0W) rules across states on the optimal NPK ratio: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variables is the distance of the consumption of N:P:K ratio from the optimal ratio of 4:2:1. The distance is calculated as follows:

$$Distance = \sqrt{(\frac{N}{N+P+K} - \frac{4}{7})^2 + (\frac{P}{N+P+K} - \frac{2}{7})^2 + (\frac{K}{N+P+K} - \frac{1}{7})^2}$$

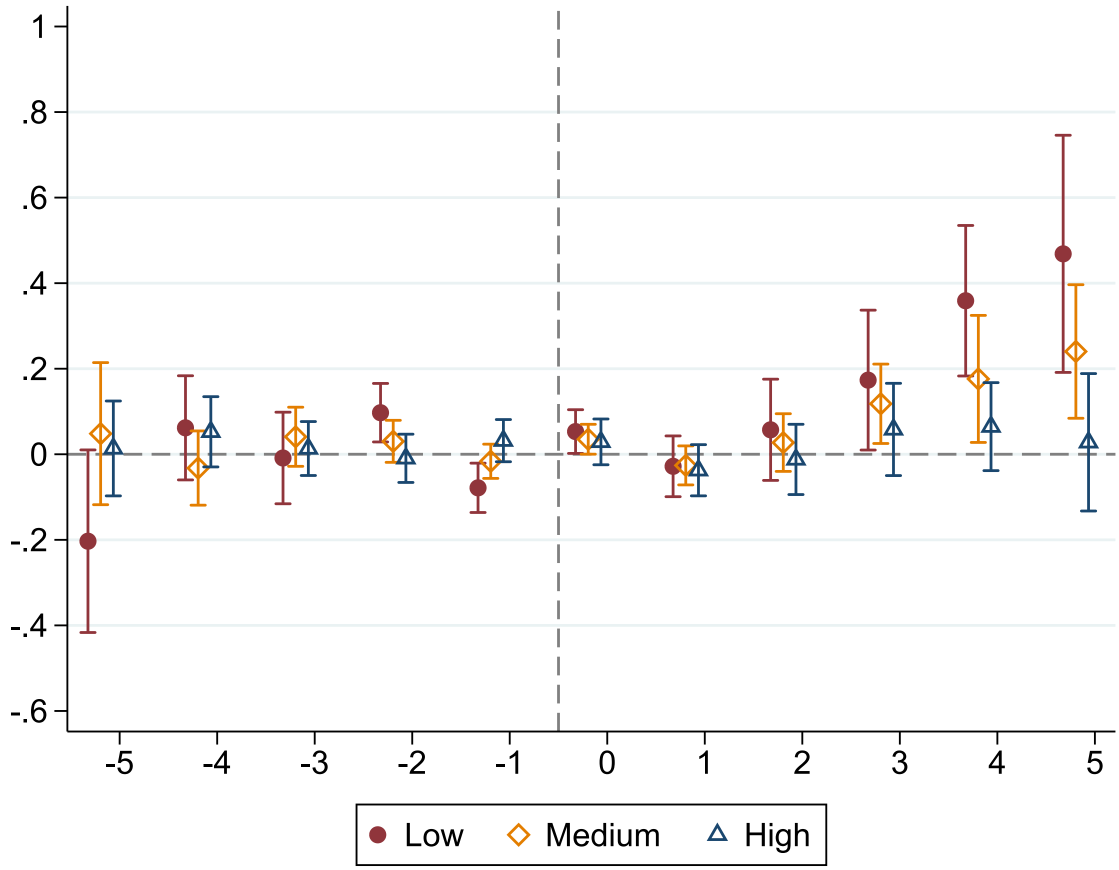
All variables are measured at the district-year level from 1998 until 2020. Treatment timing is constructed using the dates of introduction of RoW rules across different States/Union Territories as listed in C.2. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 10: 4G Introduction & effect on bank credit



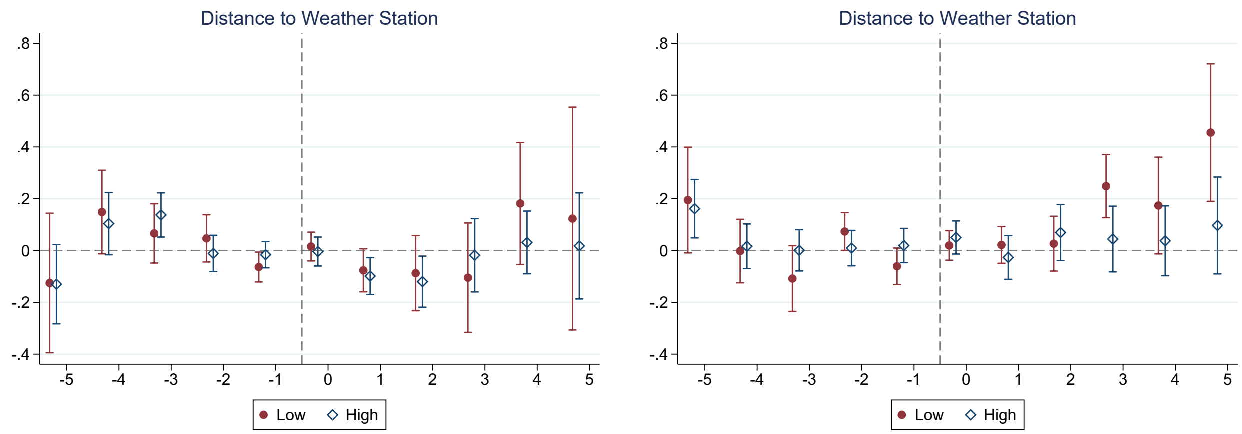
DiD estimates for impact of (staggered) passage of Rights of Way (R0W) rules across states on the growth of bank credit: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is the log of quarterly credit disbursed at the branch level. Treatment timing is constructed using the dates of introduction of RoW rules across different States/Union Territories as listed in C.2. All measures are winsorized at 1% level. Standard errors are clustered at the zipcode level. The bars represent 95 percent confidence intervals. Event-time (X-axis) is measured in quarters.

Figure 11: Heterogeneous Treatment Effect by Distance to Nearest Weather Station



DiD estimates for heterogeneous treatment effect by reliability of information (distance to nearest weather station): This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. Noise in weather information is measured using the distance of the centroid of the hexagon to the nearest weather station. We split the hexagons into three sub-samples based on the values of the distance to the nearest weather station – (1) low, is a sub-sample of hexagons with distance lower than the 25th percentile value, (2) medium, is a sub-sample of hexagons with distance greater than (or equal to) the 25th percentile value and less than (or equal to) the 75th percentile value, and (3) high, is a sub-sample of hexagons with distance greater than the 75th percentile value. **Hexagons located closer to the weather station have lower noise in information or a greater reliability of information.** All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 12: Distance to Nearest Weather Station & Rainfall Volatility

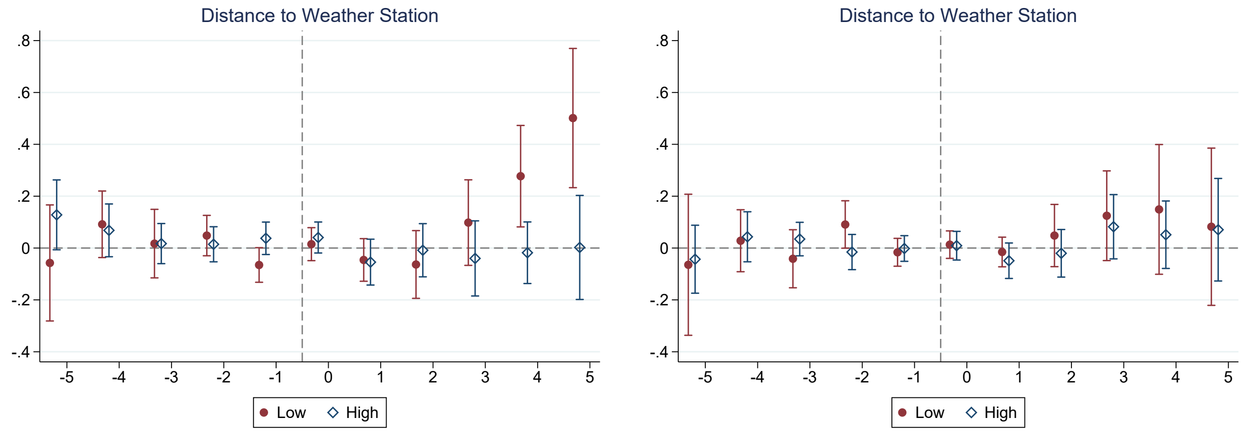


(a) Low Rainfall Volatility

(b) High Rainfall Volatility

DiD estimates for heterogeneous treatment effect by value (rainfall volatility) and reliability (distance to nearest weather station) of information: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. Reliability of weather information is measured using the distance of the centroid of the hexagon to the nearest weather station. We split the hexagons into two sub-samples based on the values of the distance to the nearest weather station – (1) low, is a sub-sample of hexagons with distance lower than the median value (blue), and (2) high, is a sub-sample of hexagons with distance greater than the median value (red). **Hexagons located closer to the weather station have lower noise in information or a greater reliability of information.** Furthermore, we define a weather station to be valuable if the history of monsoon rainfall recorded at that station exhibits low volatility. Station level rainfall volatility is computed as the coefficient of variation (CV) of rainfall using data during the monsoon from 2001 until 2012, one year before the start of the DID sample. We split stations into two buckets based on the CV value – (1) low, is a sub-sample of units (hexagons) for which the monsoon rainfall CV at the nearest weather station is lower than the median value, and (2) high, is a sub-sample of units for which the monsoon rainfall CV at the nearest weather station is higher or equal to than the median value. Panel 12a and 12b report results for hexagons with low and high volatility of rainfall. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 13: Distance to Nearest Weather Station & Predictability of Weather Shocks

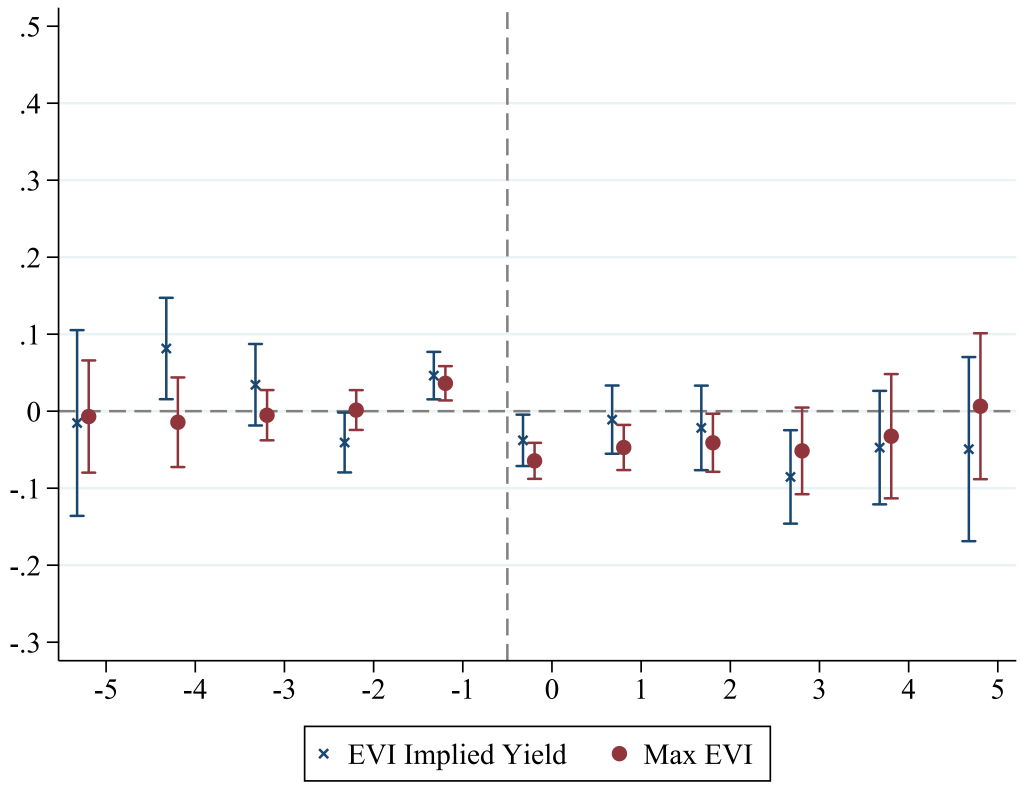


(a) Predictable Weather Shocks

(b) Difficult-to-Predict Weather Shocks

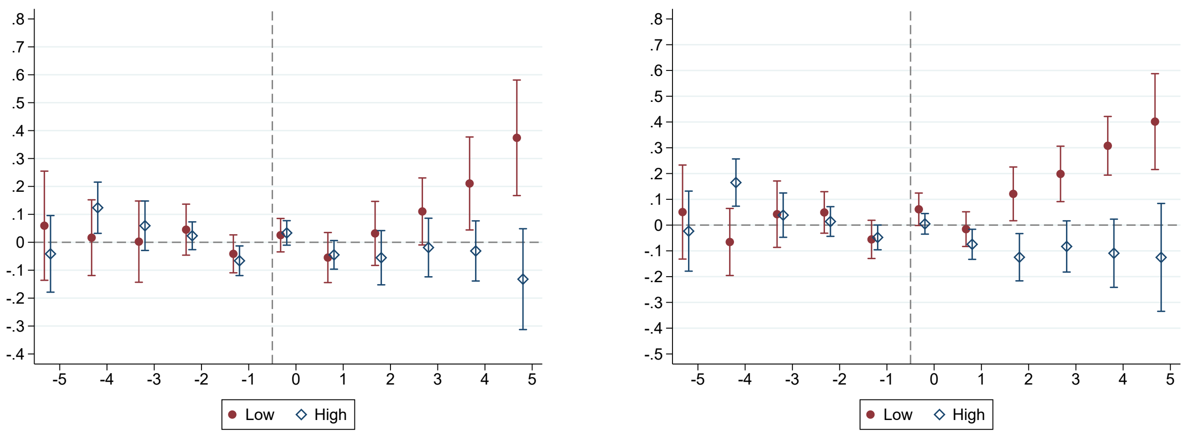
DiD estimates for heterogeneous treatment effect by relevance (predictability of weather shocks) and reliability of information (distance to nearest weather station): This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. Noise in weather information is measured using the distance of the centroid of the hexagon to the nearest weather station. We split the hexagons into two sub-samples based on the values of the distance to the nearest weather station – (1) low, is a sub-sample of hexagons with distance lower than the 50th percentile value (blue), and (3) high, is a sub-sample of hexagons with distance greater than or equal to the 50th percentile value (red). **Hexagons located closer to the weather station have lower noise in information or a greater reliability of information.** Panel 13a plots the estimates for the hexagons which are susceptible to predictable weather shocks. Predictable weather shocks include mist, drizzle and rainfall. We define a unit to be susceptible to predictable weather shock if the probability of the event occurring based on daily historical data of the occurrence from 2001 until 2012 is below the median value for the entire sample of units. Panel 13b plots the estimates for the hexagons which are susceptible to difficult to predict weather shocks. Difficult to predict weather shocks include lightning, haze, sand or dust storm, fog, squall, gale, thunderstorm, an hailstorm. We define a unit to be susceptible to difficult-to-predict weather shock if the probability of the event occurring based on daily historical data of the occurrence of these events from 2001 until 2012 is above the median value for the entire sample of units. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 14: Effect of 3G introduction on agricultural yield



DiD estimates for our baseline measures – EVI implied agricultural yield and maximum EVI: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield (in blue) constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. The second outcome variable is Max EVI (in maroon) constructed by using the maximum value of EVI during the kharif season. The dependent variables are standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 3G BTS. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure 15: Heterogeneous Treatment Effect by Trust

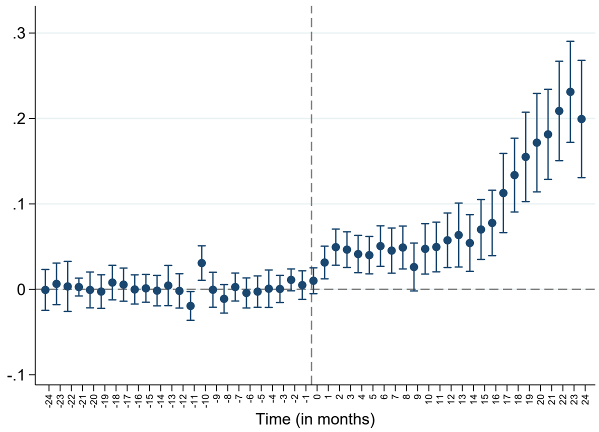


(a) Trust in Local Village Panchayat

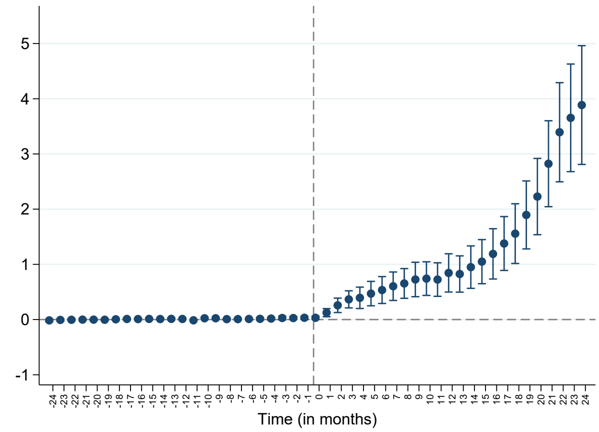
(b) Trust in State Government

DiD estimates for heterogeneous treatment effect by trust in village panchayat: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. In Figure 15a, we utilize data from India Human Development Survey (IHDS-II), conducted in 2011, which includes a question that queries respondents about the level of trust they have in their local village panchayats. The exact question is as follows: "How much trust do you place in the ability of your village panchayat to implement public projects?" The response options consist of three choices: a) A great deal of confidence, b) Only some confidence, and c) Hardly any confidence at all. To quantify these responses, we assign a numerical value of 1 to option (a), 0.33 to option (b), and 0 to option (c). We take the average of this measure across all households within a district based on their weights. This allows us to compute a continuous measure of confidence in village panchayats at the district level. We use the response to this question to construct a district-level proxy for confidence in village panchayats and divide the districts into two subsets based on the median response value. Districts with confidence levels above the median are categorized as "high confidence" districts, while those below the median are classified as "low confidence" districts. We then link this measure to the hexagonal grids situated within the districts. In Figure 15b, we utilize data from India Human Development Survey (IHDS-II), conducted in 2011, which includes a question that queries respondents about the level of trust they have in their local village panchayats. The exact question is as follows: "How much trust do you place in the ability of the state government to take care of people?" The response options consist of three choices: a) A great deal of confidence, b) Only some confidence, and c) Hardly any confidence at all. To quantify these responses, we assign a numerical value of 1 to option (a), 0.33 to option (b), and 0 to option (c). We take the average of this measure across all households within a district based on their weights. This allows us to compute a continuous measure of confidence in state government at the district level. We use the response to this question to construct a district-level proxy for confidence in state government and divide the districts into two subsets based on the median response value. Districts with confidence levels above the median are categorized as "high confidence" districts, while those below the median are classified as "low confidence" districts. We then link this measure to the hexagonal grids situated within the districts. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

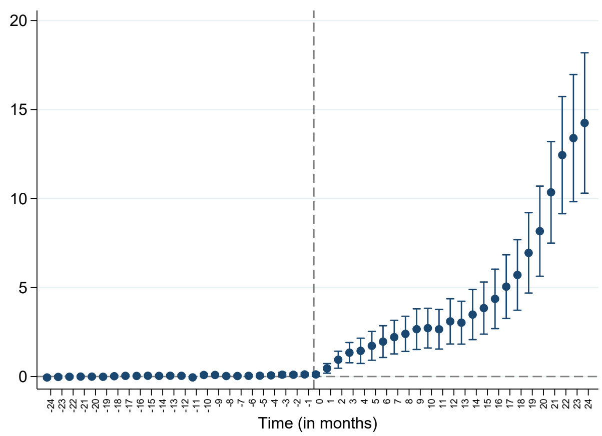
Figure 16: Effect of 4G Introduction on Krishify App Installations



(a) Extensive Margin



(b) # Installations



(c) $\frac{\#Installations}{\text{Pre-Period Average}}$

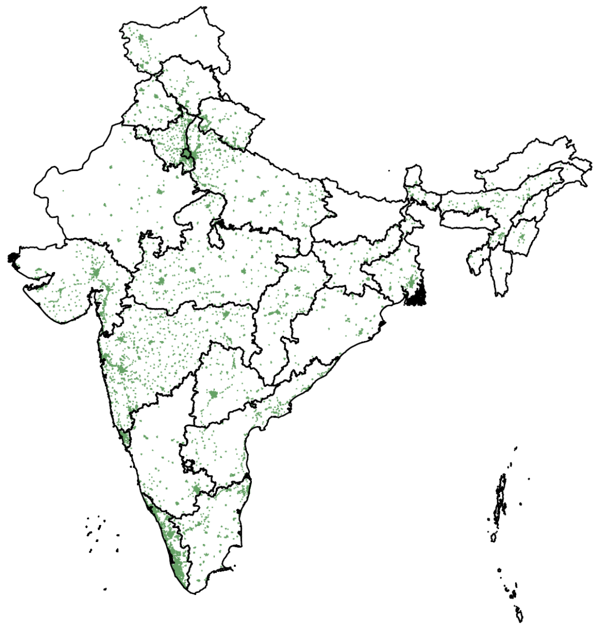
DiD estimates for impact of 4G introduction on monthly installations of Krishify mobile application at the hexagon level: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). Figure 16a uses a binary variable which takes the value of one if total monthly app downloads at the hexagon level are greater than zero, representing the extensive margin. Figure 16b uses the number of monthly installations at the hexagon level as the dependent variable. Figure 16c uses the number of monthly installations scaled by pre-period average number of monthly installations at the hexagon level as the dependent variable. Geolocations and timestamps of Krishify app downloads by users are superimposed on the hexagons in our sample to calculate monthly downloads at the hexagon level. Treatment timing in both the figures is the month of introduction of the first 4G BTS inside the boundary of a hexagon. Standard errors are clustered at the district level. The bars represent 95 percent confidence intervals.

Online Appendix for:

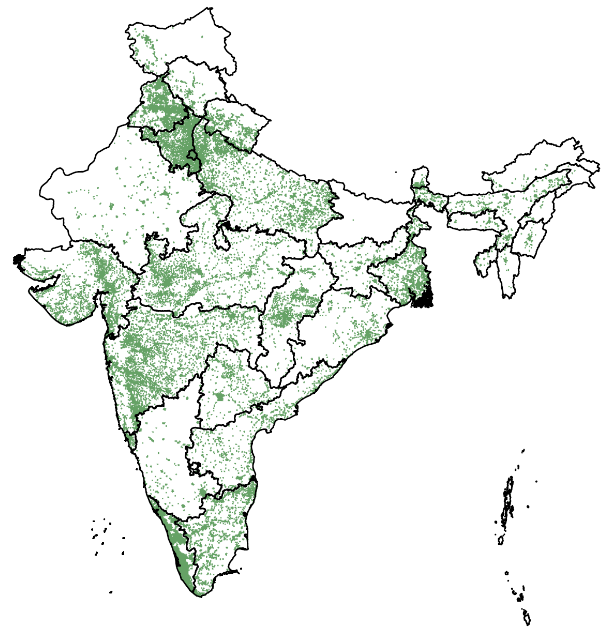
"Information Access and Local Economic Development: Evidence from the Introduction of High-Speed Internet"

Appendix A Data

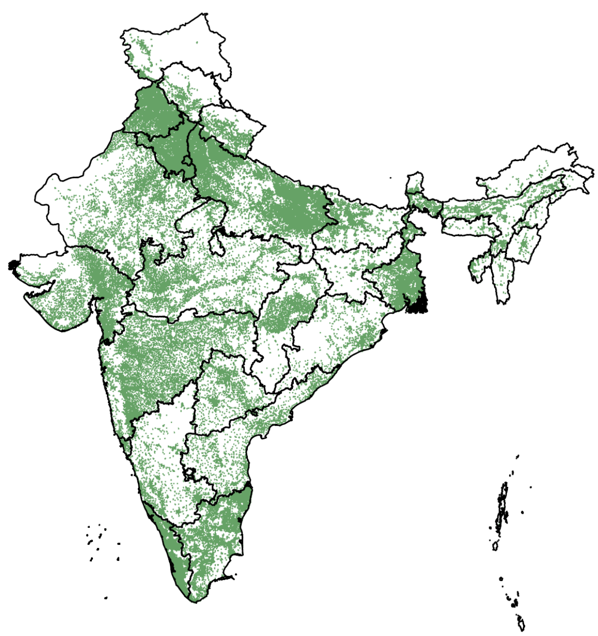
Figure A.1: Evolution of 3G coverage over time



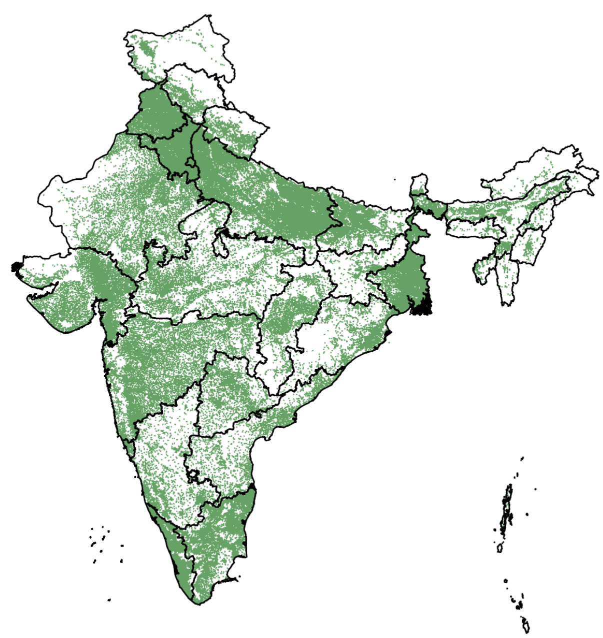
(a) 3G coverage in 2014



(b) 3G coverage in 2016



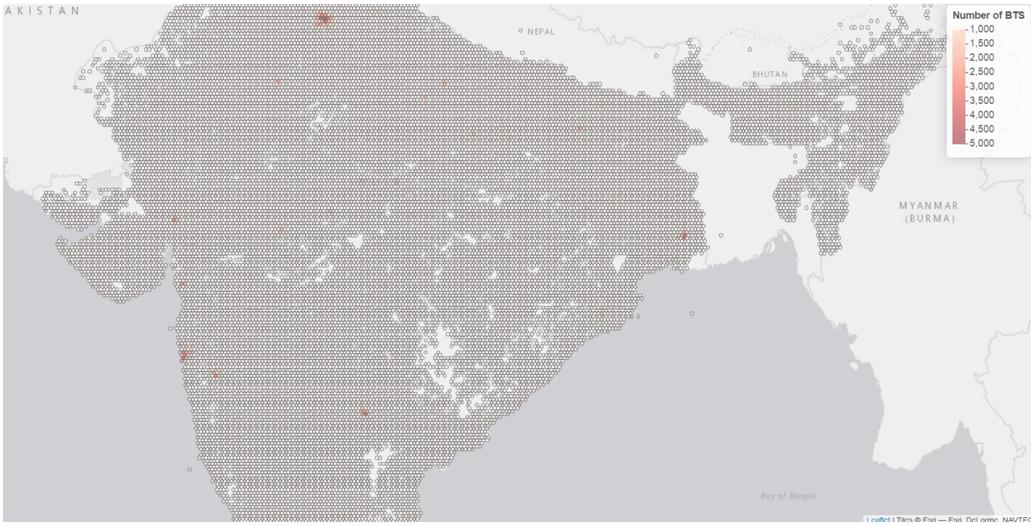
(c) 3G coverage in 2018



(d) 3G coverage in 2020

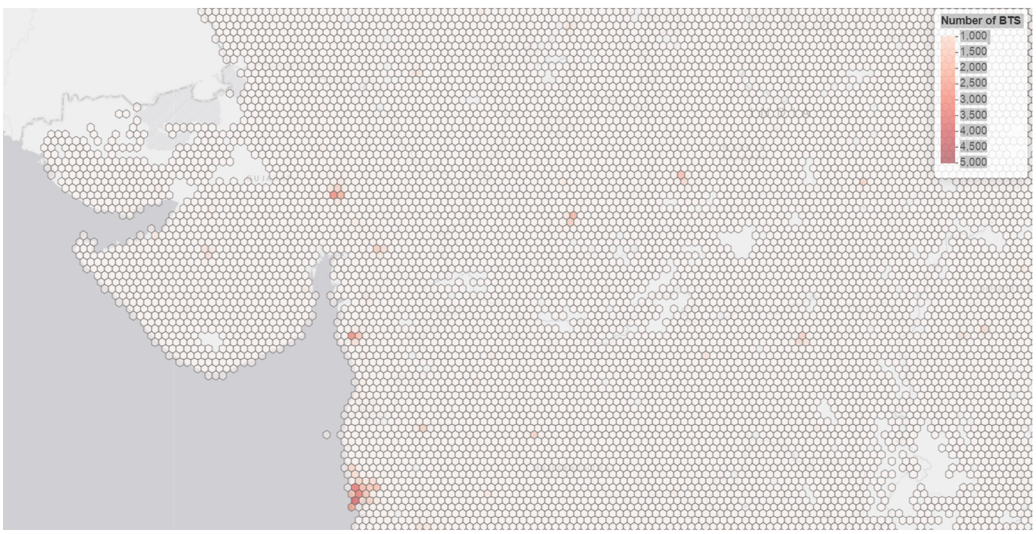
This figure plots the evolution of 3G coverage over time based on the Ministry of Telecom data.

Figure A.2: Hexagonal Tesselation of the Indian Map



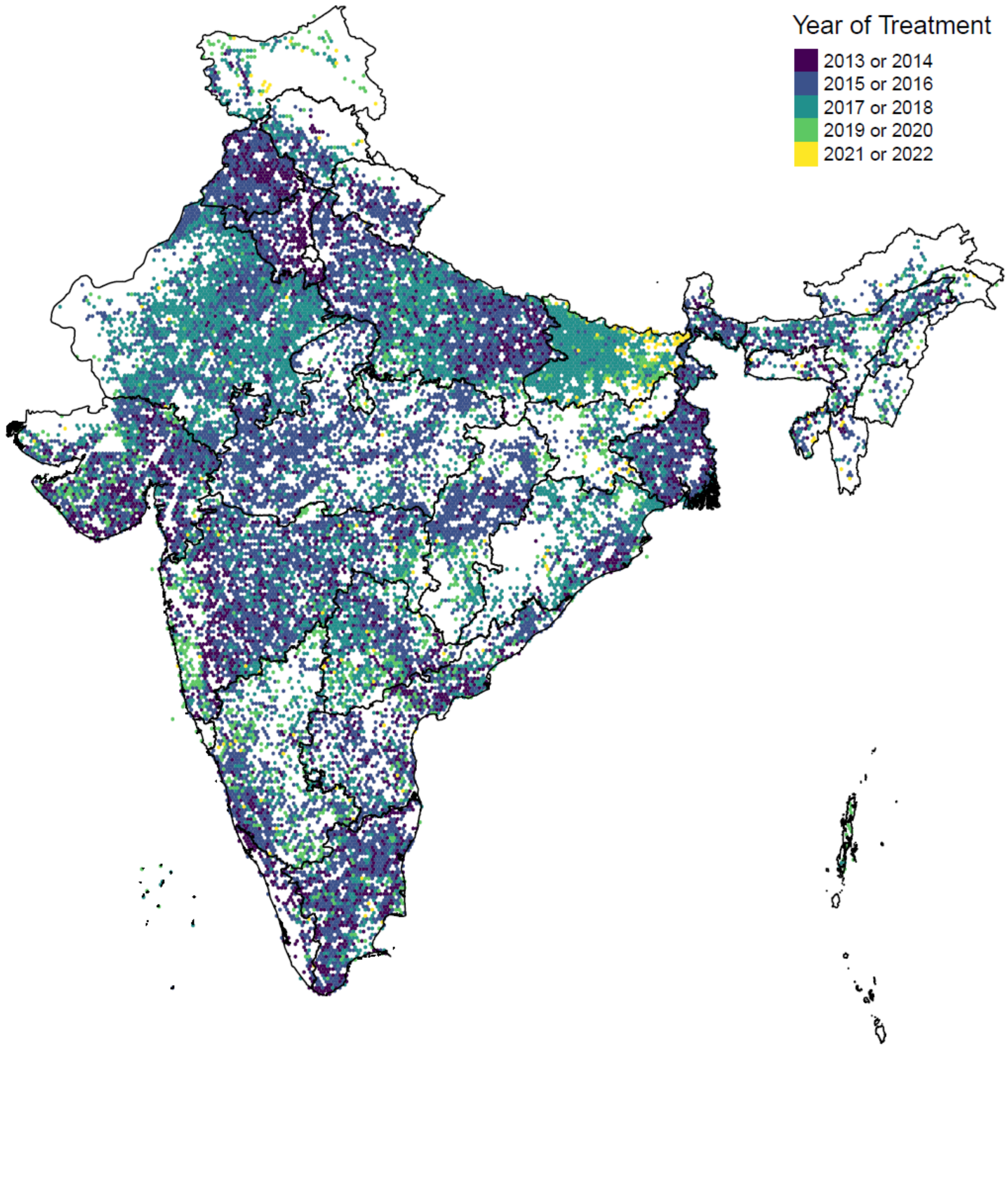
This figure shows the hexagonal units of equal area (≈ 95 sqkm.) drawn on the Indian map with at least one BTS.

Figure A.3: Hexagonal Tesselation of the Indian Map (zoomed-in)



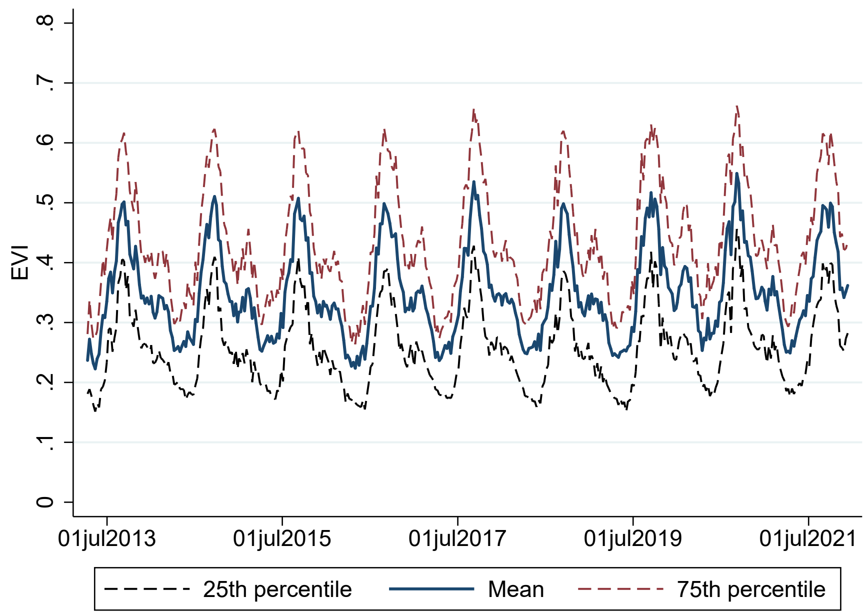
This figure shows the zoomed-in image of hexagonal units presented in Appendix Figure A.2.

Figure A.4: Evolution of 3G Treatment Across Hexagons

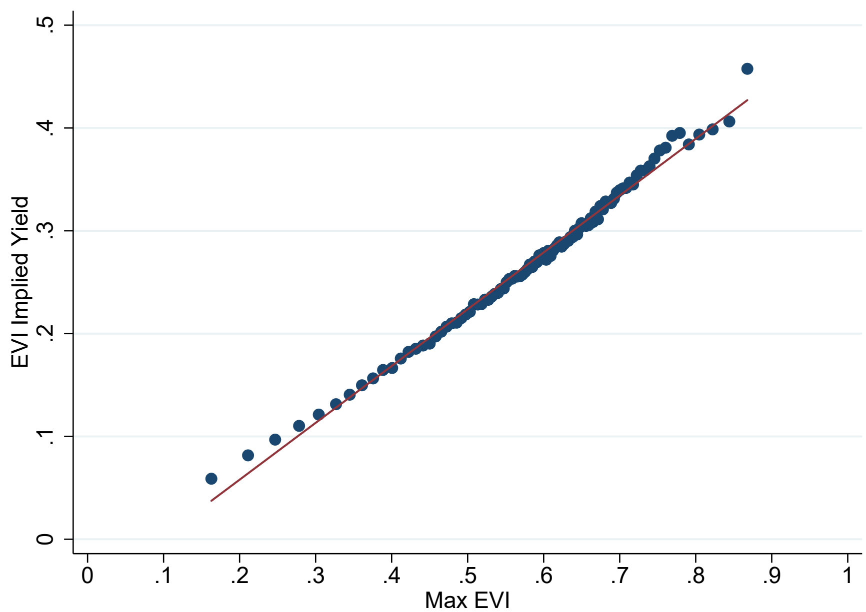


This figure presented a tessellated heat map of India. Each color denotes the year in which the 3G technology was introduced in the hexagon for the first time.

Figure A.5: Properties of Enhanced Vegetation Index (EVI)



(a) Summary Statistics of EVI



(b) Relationship between EVi Implied Yield and Maximum EVI

Appendix Figure A.5a plots the evolution of average, the 25th percentile, and the 75th percentile EVI value for *all* hexagons in our sample. Appendix Figure A.5b presents the binscatter plot of the EVI implied yield measure and the maximum EVI measure. EVI implied yield is constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. All measures are winsorized at 1% level.

Appendix B EVI Implied Yields & Agricultural Production

This section discusses the relationship between EVI implied yield and measures of agricultural production. Administrative data on actual crops yields and production values is not available beyond the district level. Therefore, we construct hexagon-level measures of actual crop yields and production values from Advancing Research on Nutrition and Agricultures (ARENA) Demographic and Health Surveys (DHS)-GIS Database. We map survey clusters in the DHS data to hexagons based on the minimum distance between the centroids of hexagons and geo-coordinates of survey clusters. Specifically, we use three measures of agricultural production available in the ARENA-DHS database. These measures include – (1) Yield of cereal crops, (2) the total value of produce of all cereal crops in USD, and (3) the total value of produce of all crops in USD. Cereal crops include a variety of different plants including wheat, rice, oats, barley, maize, rye, millet, corn, and sorghum. These measures are calculated as of the year 2014. We take a natural logarithm of these three measures.

We start by regressing these three measures against EVI-implied yields at the hexagonal level for the year 2014. Table B.1 and Figure B.1 report the regression results. Column 1 uses the natural logarithm of yield for cereal crops as the dependent variable. Yield is measured as kg of cereal crops per hectare of land. Column 2 uses the natural logarithm of total USD value of cereal crops as the dependent variable. Column 3 uses the natural logarithm of the total USD value of all crops as the dependent variable. The estimate of interest is the coefficient associated with EVI implied yield. Consistent with the findings of Asher and Novosad (2020), we observe the coefficient of interest is positive and statistically significant across all measures of agricultural production.

The size of the estimate indicates that a 0.5 unit increase in EVI implied yield is associated with a 15.1% increase in cereal yield, a 61.2% increase in USD value of cereal production, and a 45.9% increase in USD value of all crop production.

B.1 Quantifying the Effect of 4G Introduction

Next, we use these estimates to quantify the average impact of 4G introduction on actual yields and production values. Section 4 documents a 0.035 unit increase in EVI implied yield six years after the introduction of the 4G network. We use this change in EVI implied yield to impute the effect on real yields and production value. Specifically, the estimates in Table B.1 imply that the introduction of high-speed 4G internet is associated with a – (1) 1.06% increase in agricultural yields of cereal crops, (2) a 4.28% in USD production value of cereal crops, and (3) a 3.21% increase in the USD production value of all crops.

We further break down the hexagonal-level, dollar increase in production values from all crops into an estimate of impact on household-level income. Using the 2011 census data on village-level count of households, we aggregate total number of households at the hexagon-level. We do this by mapping village shapefiles to our hexagons. On average, 698.86 households reside in each hexagon. We multiply the average annual income of agricultural households in 2014 with the average number of households in each hexagon. The average monthly income for an agricultural household comes from the 2014 Situation Assessment Survey and is equal to ₹6,426. The average annual income for an agricultural household is equal to ₹77,112 (₹6,426 X 12). The average annual income for all households in a hexagon is ₹53,890,793.06 (₹77,112 X 698.86) or $ 883,464.31.

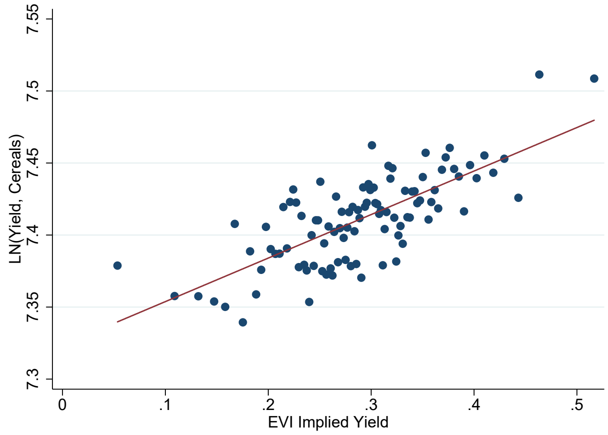
Comparing the average increase presented in columns 3 of Table B.1 with the average annual income for all households in the hexagon, we find that the introduction of 4G internet led to a 14.5% increase in household income.

Table B.1: EVI Implied Yield & Agricultural Production

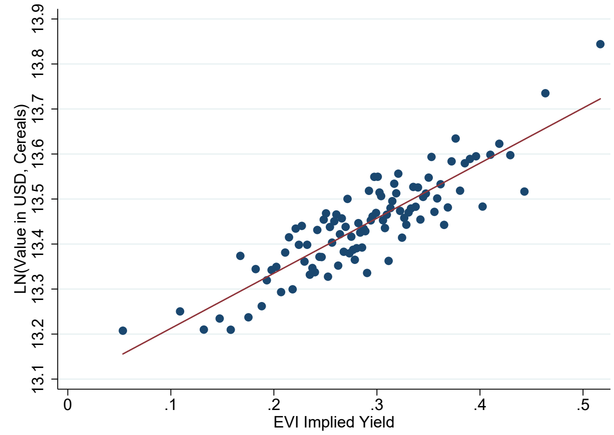
	(1)	(2)	(3)
	LN(Yield, Cereals)	LN(Value, Cereals)	LN(Value, All Crops)
EVI Implied Yield	0.3021***	1.2232***	0.9172***
	(0.0715)	(0.1854)	(0.1357)
District FE	Yes	Yes	Yes
# Obs	28,118	28,118	28,118
R^2	0.7827	0.7168	0.7104
Coef	0.3021	1.2232	0.9172
ΔX	0.035	0.035	0.035
$\frac{\Delta Y}{Y}$	1.06%	4.28%	3.21%
Mean (Y)	2,018.54 kg/hectare	$ 1,551,215.00	$ 3,989,287.00
Increase (Y)	21.34 kg/hectare	$ 66,410.62	$ 128,064.09

This table reports the relationship between EVI implied yield and three measures of agricultural production. EVI implied agricultural yield is constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. Column 1 uses the natural logarithm of yield for cereal crops as the dependent variable. Yield is measured as kg of cereal crops per hectare of land. Column 2 uses the natural logarithm of total USD value of cereal crops as the dependent variable. Column 3 uses the natural logarithm of the total USD value of all crops as the dependent variable. Cereal crops include a variety of different plants including wheat, rice, oats, barley, maize, rye, millet, corn, and sorghum. All variables are measured at the hexagon level for the year 2014. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. All variables are winsorized at 1%. Figure B.1 reports the same regressions in graphical format. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

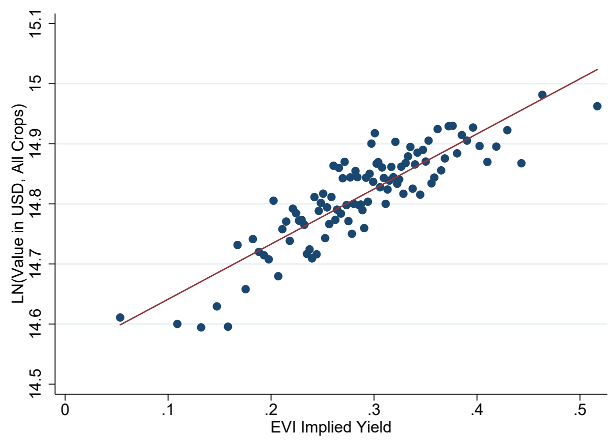
Figure B.1: EVI Implied Yield & Agricultural Production



(a) Yield for Cereals (kg per hectare)



(b) Value of Cereals Produced in USD

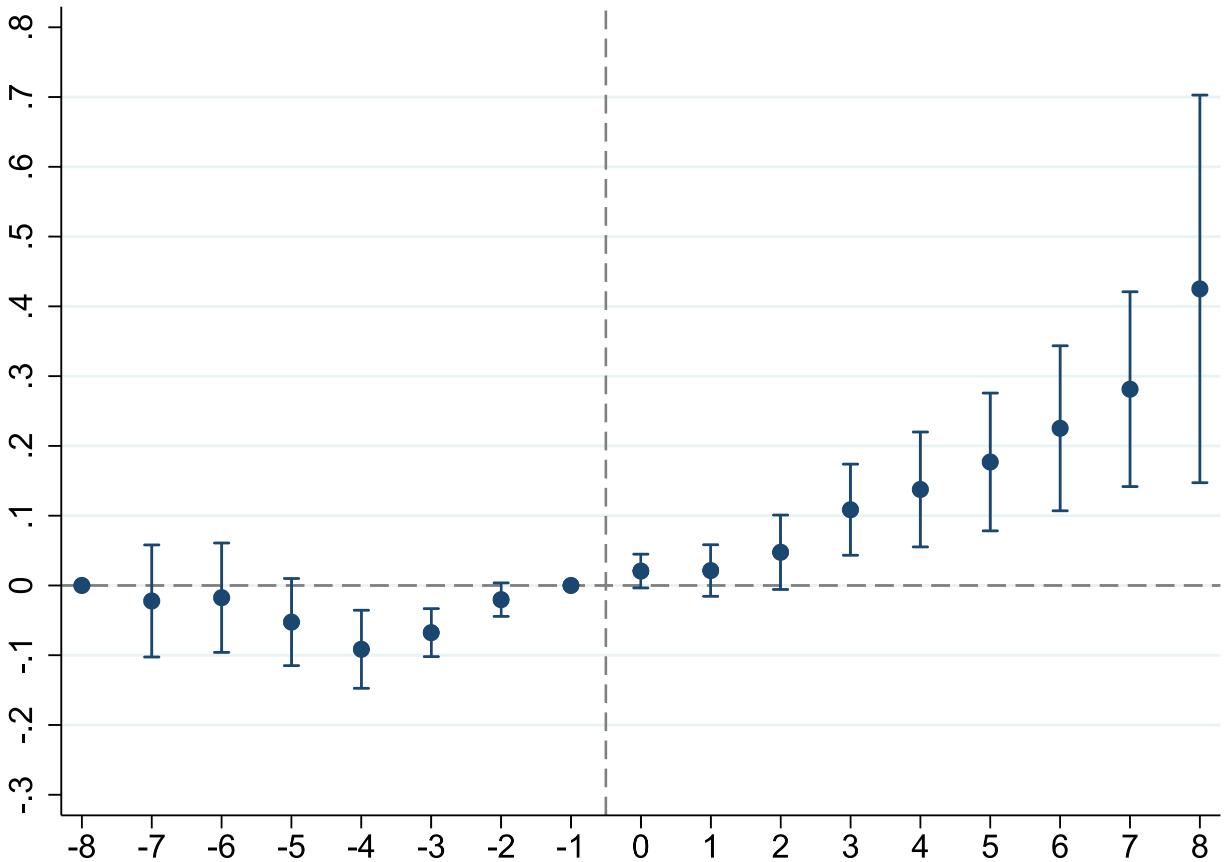


(c) Value of All Crops Produced in USD

This figure plots the relationship between EVI implied yield and three measures of agricultural production. Panel B.1a plots the relationship with the natural logarithm of yield on cereal crops measured in kg per hectare. Panel B.1b plots the relationship with the natural logarithm of the total USD value of cereal crops. Cereal crops include a variety of different plants including wheat, rice, oats, barley, maize, rye, millet, corn, and sorghum. Panel B.1c plots the relationship with the natural logarithm of total USD value of all crops. All variables are winsorized at 1%. All plots include district fixed effects.

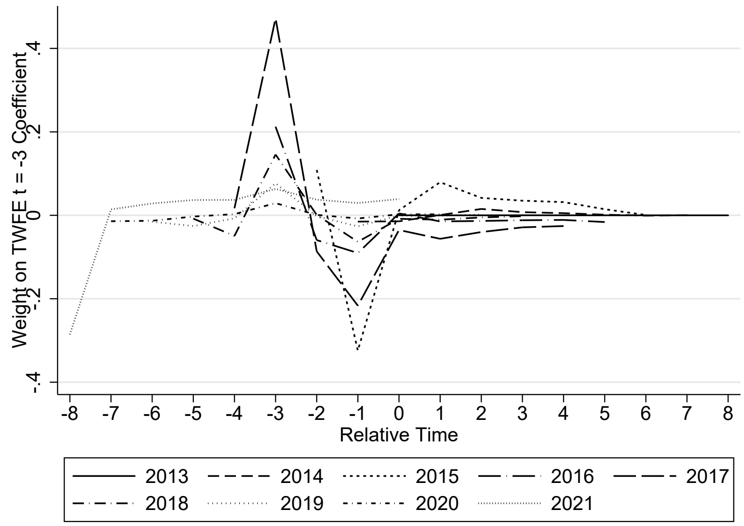
Appendix C Robustness

Figure C.1: Dynamic TWFE specification for baseline EVI measure

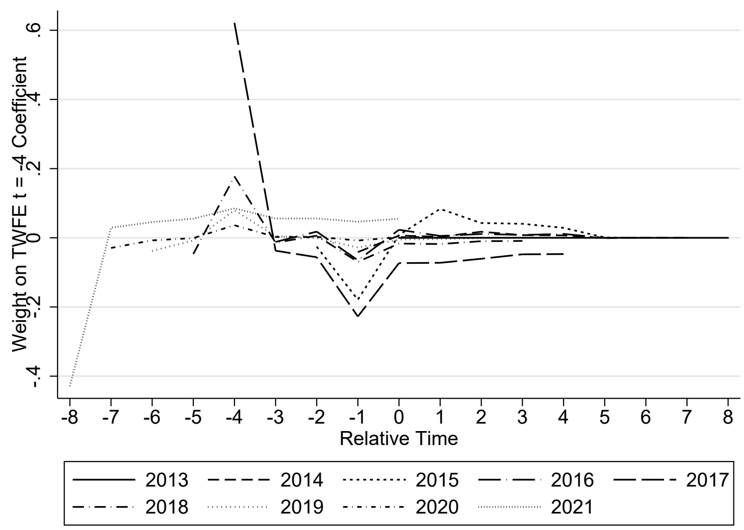


This figure plots the event-study coefficients estimated using the dynamic TWFE specification in Equation 1. The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The dependent variable is standardized. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. Excluded time periods are $t = -8, -1$. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure C.2: Weight underlying pre-treatment coefficients β_{-3} and β_{-4}



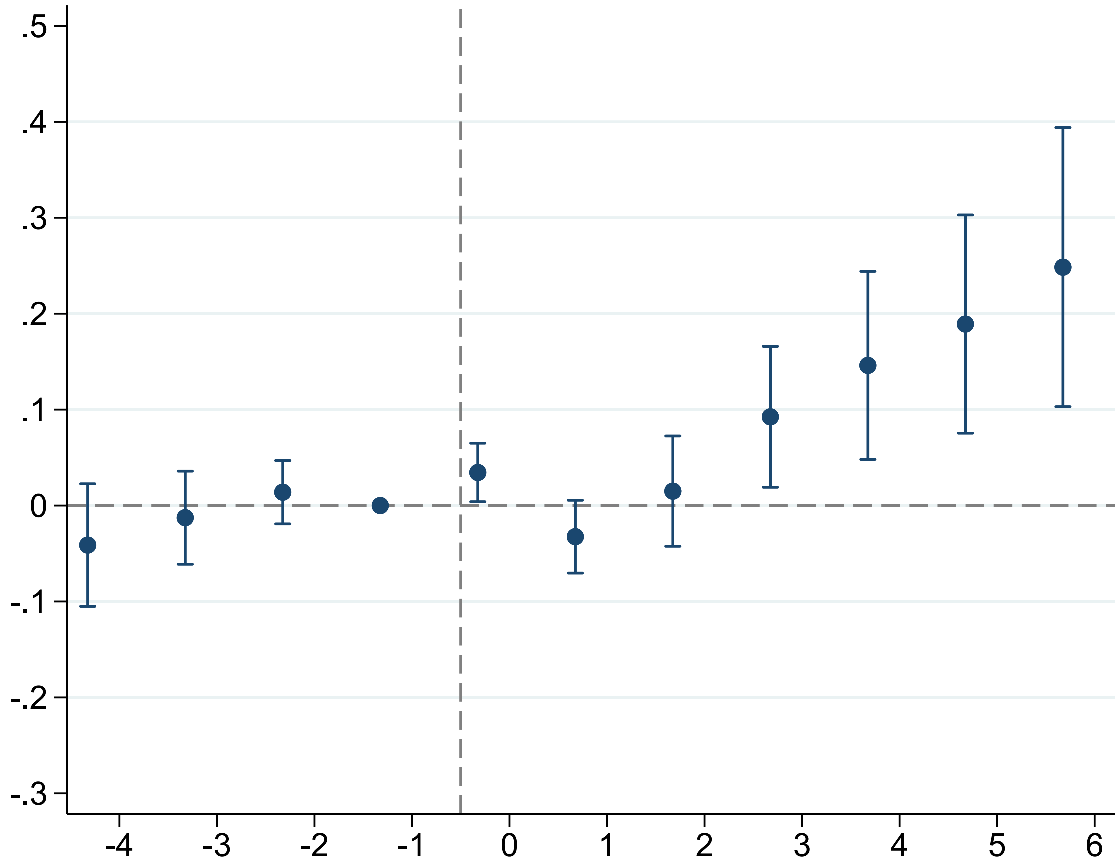
(a) Weights underlying β_{-3}



(b) Weights underlying β_{-4}

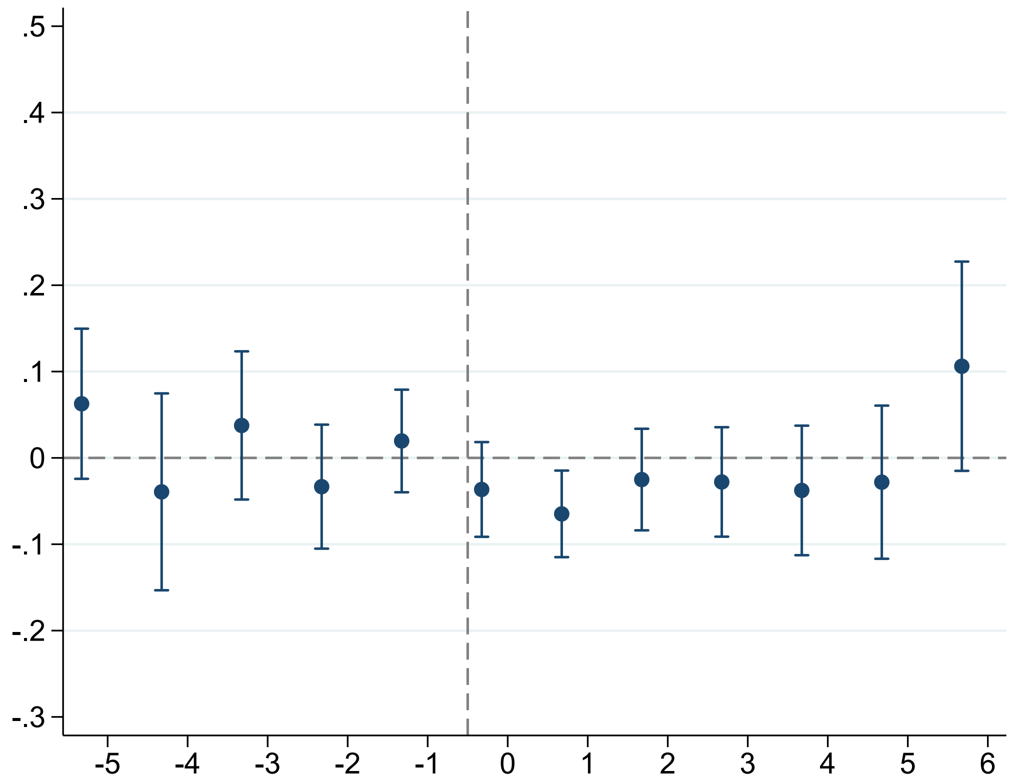
This figure plots the magnitude of weights associated with each $CATT_{e,l}$ in the weights decomposition of β_{-3} and β_{-4} as specified in Equation 2. $CATT_{e,l}$ refers to average treatment effect for cohort e in relative time period l. For instance, $CATT_{2015,-2}$ is the average treatment effect for the relative time period $t = -2$ for the cohort that got treated in the year 2015. The weights are estimated using the methodology outlined in Sun and Abraham (2021). Figures C.2a and C.2b plot the weights underlying β_{-3} and β_{-4}, respectively.

Figure C.3: Callaway and Sant'Anna (2021) baseline EVI measure with pre-treatment effects as long differences



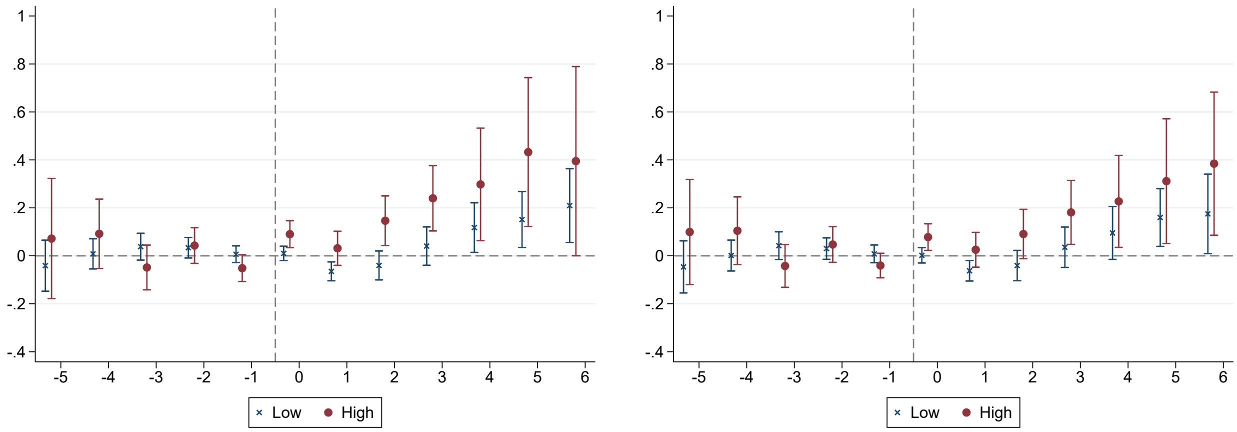
This figure plots the event-study dynamic coefficients estimated using the Callaway and Sant'Anna (2021) estimator with pre-treatment effects calculated as long differences. As pointed out in Roth (2024), pre-treatment effects in their default implementation in Stata do not match those of the traditional TWFE event-study regressions. Using the `long2` option in the Stata implementation of the Callaway and Sant'Anna (2021) estimator, they can be made comparable to conventional TWFE plots. The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of Kharif season from the maximum value of EVI during the Kharif season. The dependent variable is standardized. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure C.4: Robustness: Falsification using sample of hexagons with no cropland



This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is the maximum EVI of the kharif season. The dependent variable is standardized to mean zero and standard deviation of one. We use the maximum EVI measure here instead of EVI implied yield. We do not use EVI implied yield because it is a difference measure and is mostly centered around zero for non-croplands. This is because, unlike the cropped areas, the measure does not exhibit significant variation across the kharif season. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021 and have zero cropland as of 2010 according to the GFSAD database. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

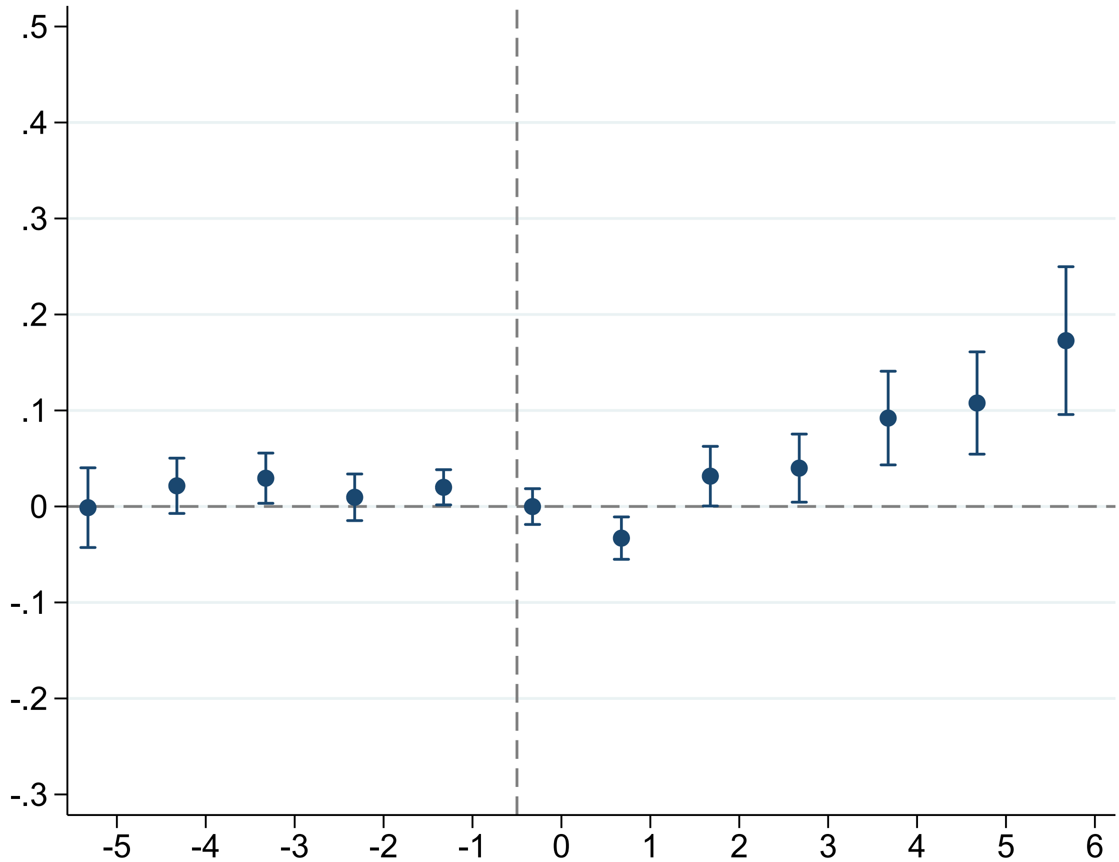
Figure C.5: Heterogeneous Treatment Effect by Agricultural Production Potential



(a) Available Land-based measure of production potential (b) Global Land-based measure of production potential

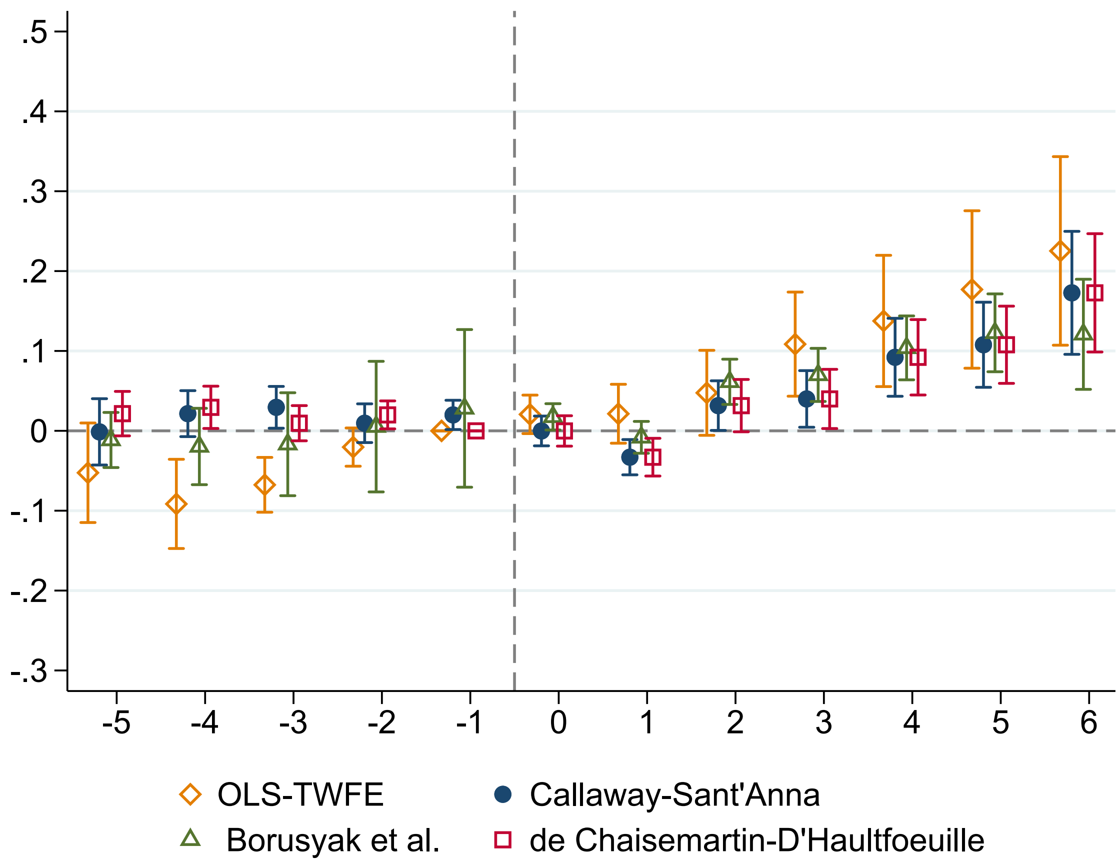
This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. The dependent variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4G BTS between 2013 and 2021. Agricultural production potential is measured using two measures provided in Advancing Research on Nutrition and Agriculture's (ARENA) Demographic and Health Surveys (DHS)-GIS Database. Measure 1 provides the combined suitability of currently available land for pasture and rainfed crops. Measure 2 provides the combined suitability of the global land area for pasture and rainfed crops. We split the hexagons into two sub-samples based on the values of agricultural production potential. A hexagon is defined as an area with high agricultural production potential if it is marked as land well-suited or prime land for rainfed crops, i.e., the CSI (Crop Suitability Index) is greater than equal to 50; otherwise, the hexagon is classified as low potential area. Panel C.5a and C.5b report result for measures 1 and 2 based on total available land and total global land, respectively. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure C.6: Alternative Measure: Effect on maximum EVI



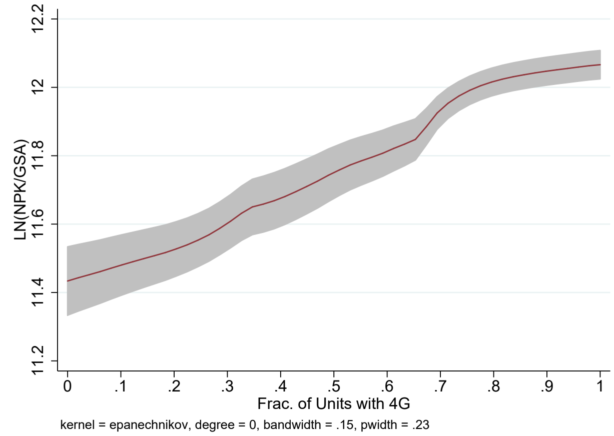
DiD estimates for alternative measure – maximum EVI during the season: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is maximum EVI during the Kharif season. The sample consists of all unique hexagons that saw the introduction of 4g BTS between 2014 and 2022. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure C.7: Robustness using alternative DiD estimators: Maximum EVI

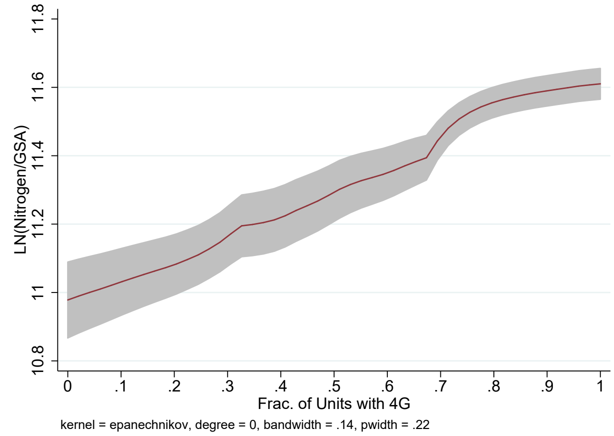


Robustness with multiple DiD estimators for alternative measure – maximum EVI during the season This figure plots the event-study dynamic coefficients using four different estimators: (1) dynamic TWFE estimator (in orange with hollow diamond markers), (2) Callaway and Sant'Anna (2021) estimator (in blue with solid circle markers), (3) Borusyak, Jaravel and Spiess (2022) estimator (in green with hollow triangle markers), and (4) De Chaisemartin and d'Haultfoeuille (2020) estimator (in pink with hollow square markers). The outcome variable is maximum EVI during the Kharif season. The outcome variable is standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 4g BTS between 2013 and 2021. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

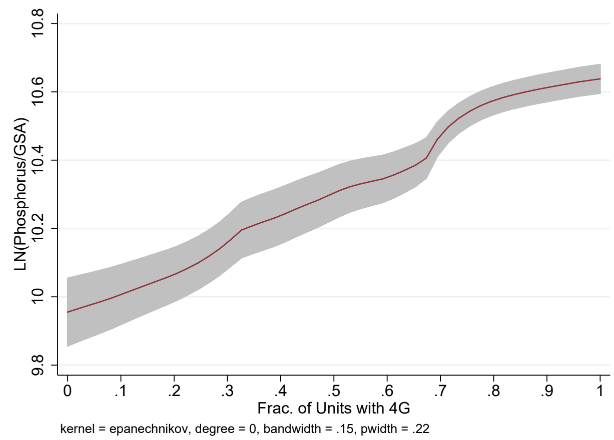
Figure C.8: 4G introduction & Fertilizer Usage



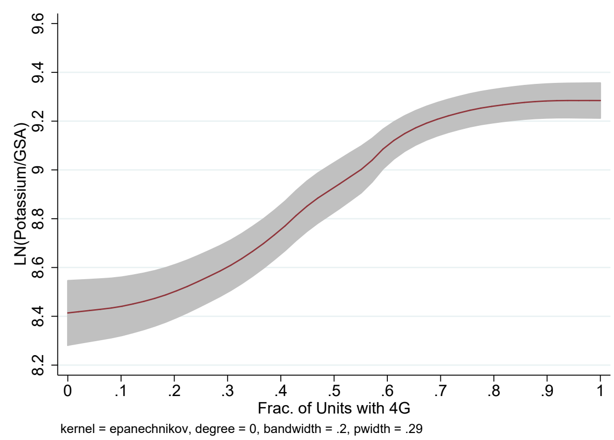
(a) Total Fertilizer Usage (NPK)



(b) Nitrogen Usage



(c) Phosphate Usage



(d) Potash Usage

4G penetration on fertilizer usage: This figure presents the local polynomial plot of fertilizer usage and 4G penetration at the district-level. Fertilizer consumption is measured as the natural logarithm of the amount of consumption of total fertilizers (NPK), nitrogen (N), phosphate (P), and potash (K) per unit of gross sown area (GSA). 4G penetration is measured as the fraction of hexagons within the district that that at least one 4G tower during the year. All variables are measured at the district-year level from 2014 until 2020. The solid red line denotes the best fit line. The gray shaded region denotes the 95% confidence intervals. The different parametric choices related to kernel, degree, bandwidth and pwidth are reported next to each figure. All variables are winsorized at 1%.

Table C.1: 4G Penetration & Fertilizer Consumption

	(1)	(2)	(3)	(4)
	LN(NPK/GSA)	LN(N/GSA)	LN(P/GSA)	LN(K/GSA)
4G penetration	0.1801***	0.2061***	0.1464**	0.2614**
	(0.0576)	(0.0623)	(0.0676)	(0.1102)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# Obs	2,585	2,585	2,585	2,585
R^2	0.9563	0.9563	0.9249	0.9247

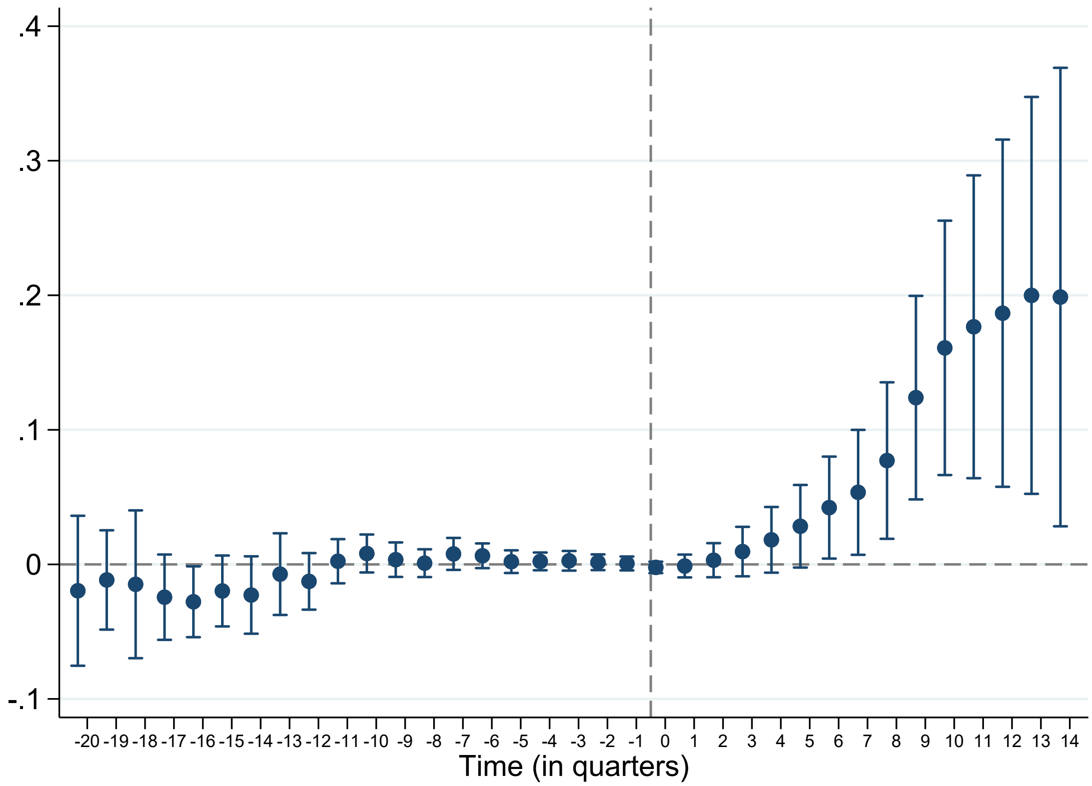
This table reports the relationship between 4G penetration and fertilizer consumption. Fertilizer consumption is measured as the natural logarithm of the amount of consumption of total fertilizers (NPK), nitrogen (N), phosphorus (P), and potassium (K) per unit of gross sown area (GSA). 4G penetration is measured as the fraction of hexagons within the district that that at least one 4G tower during the year. All variables are measured at the district-year level from 2014 until 2020. All variables are winsorized at 1%. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.2: Rights of Way (RoW) rules status across states

	State	RoW policy notified/ approved by Cabinet	Approval Date	Draft Policy released	Existing policy under advance discussion	No uniform policy
1	Jharkhand	Yes	2015-12-04	-	-	-
2	Rajasthan	Yes	2017-02-06	-	-	-
3	Tripura	Yes	2017-09-06	-	-	-
4	Odisha	Yes	2017-09-14	-	-	-
5	Haryana	Yes	2017-10-06	-	-	-
6	Assam	Yes	2018-02-16	-	-	-
7	Maharashtra	Yes	2018-02-17	-	-	-
8	Tamil Nadu	Yes	2018-02-18	-	-	-
9	Arunachal Pradesh	Yes	2018-05-10	-	-	-
10	Uttar Pradesh	Yes	2018-06-15	-	-	-
11	Uttarakhand	Yes	2018-09-13	-	-	-
12	Meghalaya	Yes	2018-12-20	-	-	-
13	Madhya Pradesh	Yes	2019-03-08	-	-	-
14	Karnataka	Yes	2019-05-29	-	-	-
15	Manipur	Yes	2019-11-28	-	-	-
16	Nagaland	Yes	2019-12-02	-	-	-
17	Gujarat	No	-	-	Yes	-
18	Daman and Diu	No	-	-	-	Yes
19	Lakshadweep	No	-	-	-	Yes
20	Dadra and Nagar Haveli	No	-	-	-	Yes
21	Goa	No	-	-	Yes	-
22	Jammu and Kashmir	No	-	Yes	-	-
23	Punjab	No	-	Yes	-	-
24	Kerala	No	-	Yes	-	-
25	Puducherry	No	-	-	Yes	-
26	Himachal Pradesh	No	-	-	-	-
27	Andhra Pradesh	No	-	-	Yes	-
28	Chandigarh	No	-	-	Yes	-
29	NCT of Delhi	No	-	Yes	-	-
30	Telangana	No	-	-	Yes	-
31	Chhattisgarh	No	-	-	Yes	-
32	Bihar	No	-	-	Yes	-
33	West Bengal	No	-	-	-	Yes
34	Sikkim	No	-	Yes	-	-
35	Mizoram	No	-	Yes	-	-
36	Andaman and Nicobar	No	-	-	-	Yes

This table lists the status of RoW policies (as of September 2020) across States/Union territories in India as per the GSMA (2020) report. The report highlights the importance of States' adoption of Rights of Way rules in line with the Indian Telegraph RoW Rules, 2016, issued by the Ministry of Telecom, Government of India. These policies were rolled out to promote the expansion of telecom infrastructure. According to the report, as of September 2020, 16 states had adopted RoW rules. Additionally, 15 states and union territories were deliberating on the passage of these rules in the state legislature.

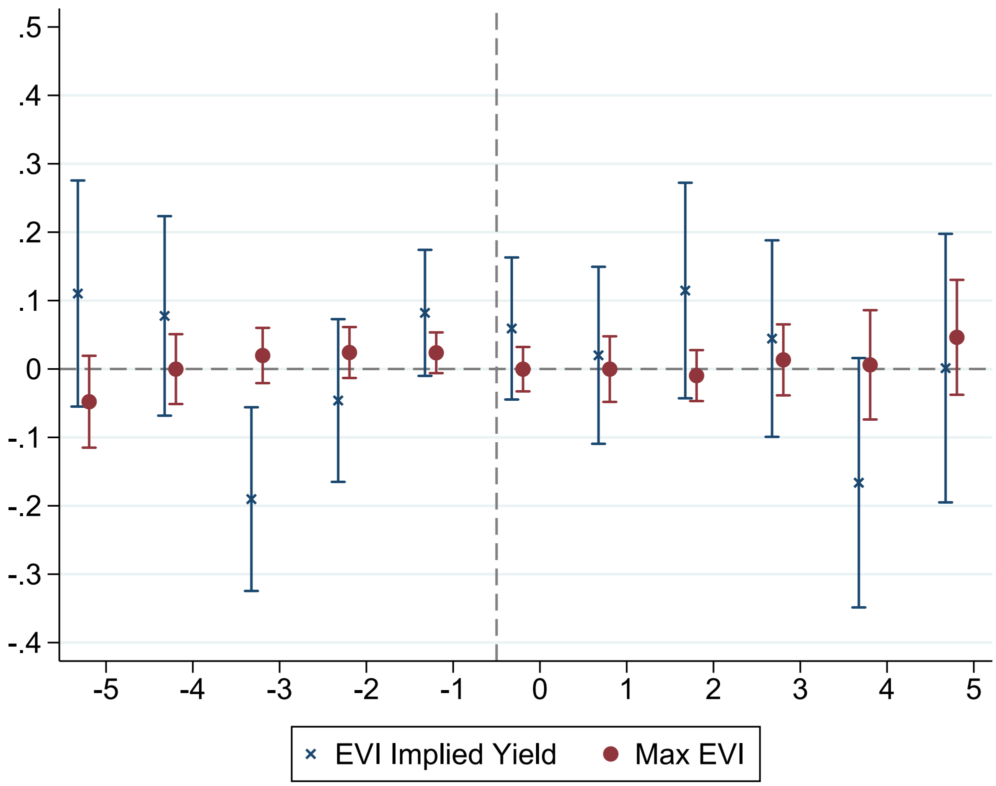
Figure C.9: Effect of 4G introduction on agricultural credit at ZIP code level



DiD estimates for impact of (staggered) introduction of 4G BTS on the growth of bank credit at the zipcode level: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is the log of quarterly agricultural credit disbursed at the zipcode level. The sample consists of all the zipcodes that saw the introduction of a 4G BTS between 2014 to 2022. Treatment timing is defined as the date on which the first 4G BTS became operational within a zipcode. Natural log of zipcode area is added as a control. All measures are winsorized at 1% level. Standard errors are clustered at the district level. The bars represent 95 percent confidence intervals. Event-time (X-axis) is measured in quarters.

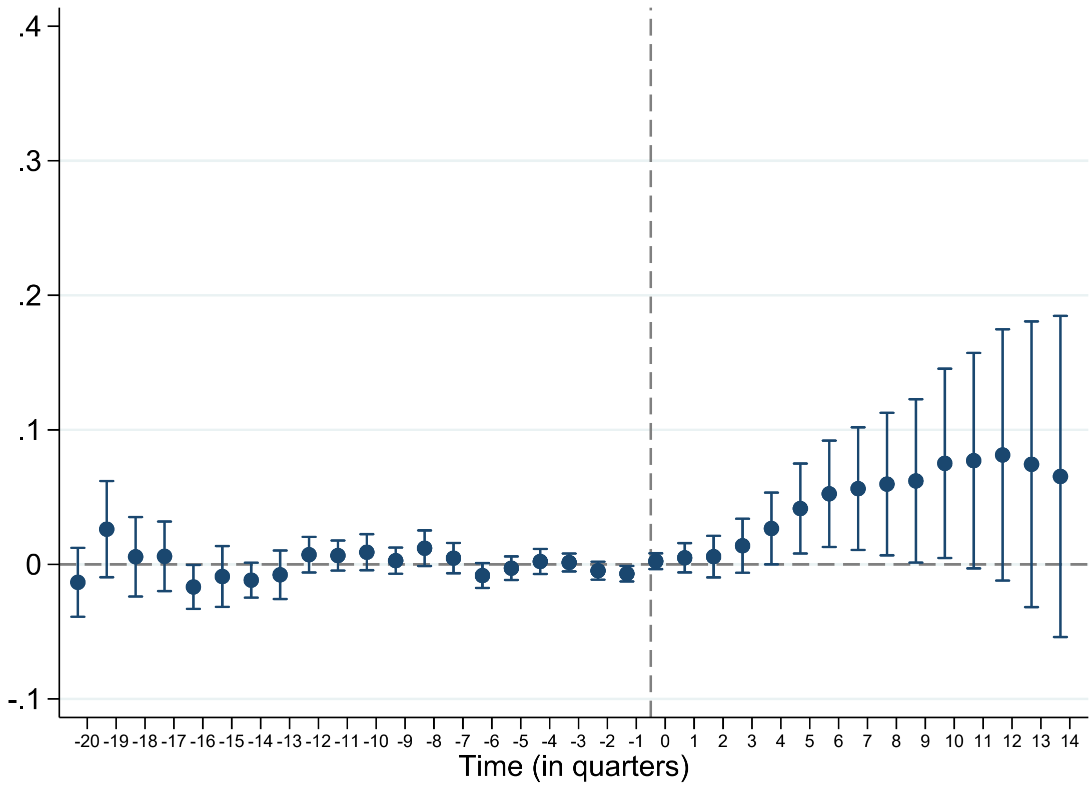
Appendix D Technology Adoption

Figure D.1: Effect of 2G introduction on agricultural yield



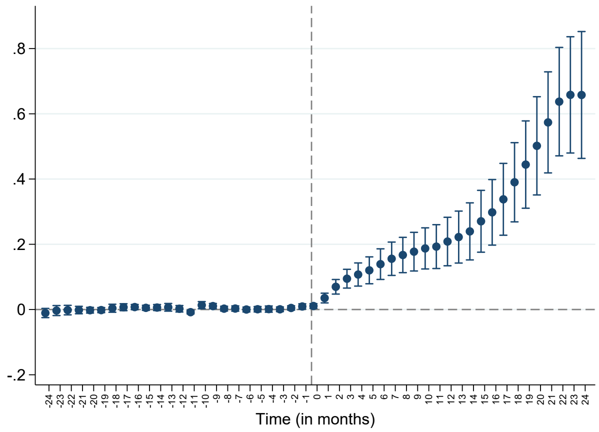
DiD estimates for our baseline measures – EVI implied agricultural yield and maximum EVI: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is EVI implied agricultural yield (in blue) constructed by subtracting the average value of EVI during the initial weeks of kharif season from the maximum value of EVI during the kharif season. The second outcome variable is Max EVI (in maroon) constructed by using the maximum value of EVI during the kharif season. The dependent variables are standardized to mean zero and standard deviation of one. The sample consists of all unique hexagons that saw the introduction of 2G BTS. All measures are winsorized at 1% level. Standard errors for all estimators are clustered at the district level. The bars represent 95 percent confidence intervals.

Figure D.2: Effect of 3G introduction on agricultural credit at ZIP code level

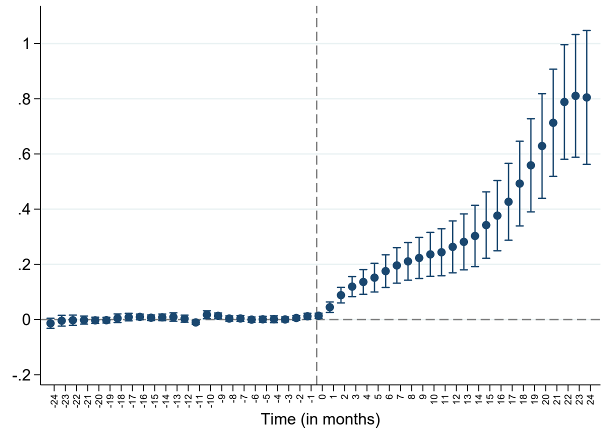


DiD estimates for impact of (staggered) introduction of 3G BTS on the growth of bank credit at the zipcode level: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). The outcome variable is the log of quarterly agricultural credit disbursed at the zipcode level. The sample consists of all the zipcodes that saw the introduction of a 3G BTS between 2014 to 2022. Treatment timing is defined as the date on which the first 3G BTS became operational within a zipcode. Natural log of zipcode area is added as a control. All measures are winsorized at 1% level. Standard errors are clustered at the district level. The bars represent 95 percent confidence intervals. Event-time (X-axis) is measured in quarters.

Figure D.3: Effect of 4G Introduction on Krishify App Installations



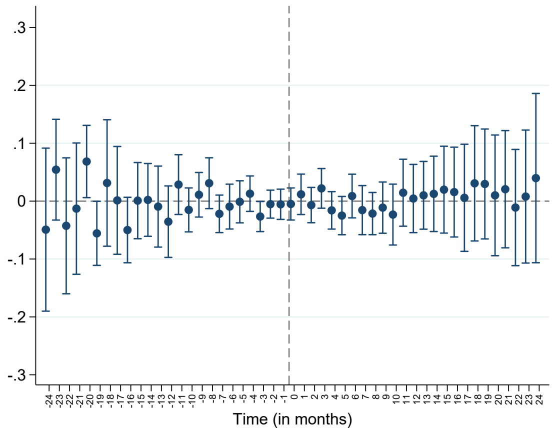
(a) Log(1+#Installations)



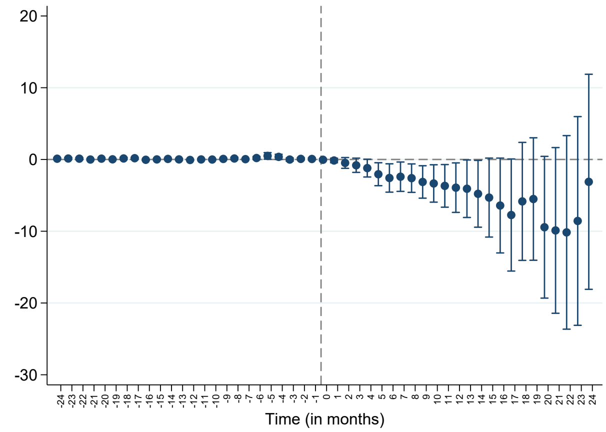
(b) IHS(#Installations)

DiD estimates for impact of 4G introduction on monthly installations of Krishify mobile application at the hexagon level: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). Figure D.3a uses the natural logarithm of the number of monthly installations at the hexagon level, respectively, as the dependent variable. Figure D.3b uses the inverse hyperbolic transformation of the number of monthly installations at the hexagon level as the dependent variable. Geolocations and timestamps of Krishify app downloads by users are superimposed on the hexagons in our sample to calculate monthly downloads at the hexagon level. Treatment timing in both the figures is the month of introduction of the first 4G BTS inside the boundary of a hexagon. Standard errors are clustered at the district level. The bars represent 95 percent confidence intervals.

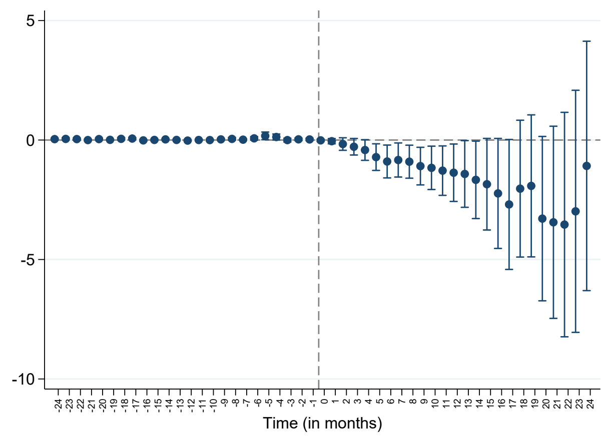
Figure D.4: Effect of 3G Introduction on Krishify App Installations



(a) Extensive Margin



(b) # Installations



(c) $\frac{\#Installations}{\text{Pre-Period Average}}$

DiD estimates for impact of 3G introduction on monthly installations of Krishify mobile application at the hexagon level: This figure plots the event-study dynamic coefficients estimated using the methodology outlined in Callaway and Sant'Anna (2021). Figure 16a uses a binary variable which takes the value of one if total monthly app downloads at the hexagon level are greater than zero, representing the extensive margin. Figure 16b uses the number of monthly installations at the hexagon level as the dependent variable. Figure 16c uses the number of monthly installations scaled by pre-period average number of monthly installations at the hexagon level as the dependent variable. Geolocations and timestamps of Krishify app downloads by users are superimposed on the hexagons in our sample to calculate monthly downloads at the hexagon level. Treatment timing in both the figures is the month of introduction of the first 3G BTS inside the boundary of a hexagon. Standard errors are clustered at the district level. The bars represent 95 percent confidence intervals.

D.1 Description of Krishify Data

Krishify is an Indian company aimed at connecting farmers on a social network where they can discuss agriculture-related issues. Link to their official website can be found here.

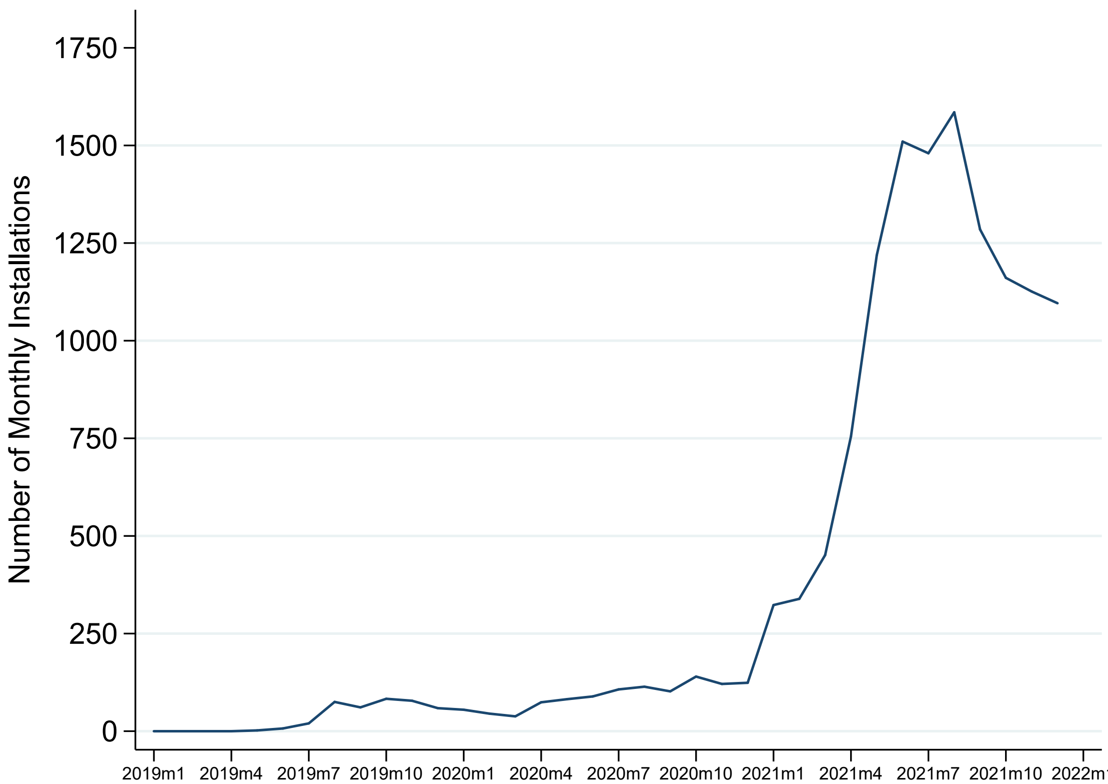
Krishify is a mobile application that can be downloaded through the Google play store with a "vision to democratize the flow of information and scale up economic opportunities for every farmer in this country". It provides a social network for farmers to connect with each other and a commerce platform for agri-businesses to connect with a community of more than 10 million farmers. They help businesses target farmers on their network mainly through engaging and informative videos. Krishify offers subscription plans to businesses to develop content for them as well as incentives to farmers and other content creators to upload videos by allowing them to monetize their content. They have more than 3.1 million daily video views, 140,000 daily active users, and 500,000 daily business interactions between farmers and businesses.

We obtain two datasets from Krishify. First, we obtain a proprietary, geolocated and time-stamped dataset of their app installations from 2019 until 2021. Second, we also obtain a 10% random sample of farmers in the Krishify database along with their detailed search, browsing, like, and comment history.

We superimpose the coordinates of app installations over our hexagonal grid to identify the number of app installations at the hexagonal level. We use this dataset to conduct a differences-in-differences analysis that examines the effect of 4G introduction on internet adoption. Therefore, we only keep hexagons where 4G was introduced after January 2019. Appendix Figure D.5 presents the evolution of monthly app installations in our sample. Appendix Figure D.6 presents the geographic distribution of total Krishify app installations by December 2021.

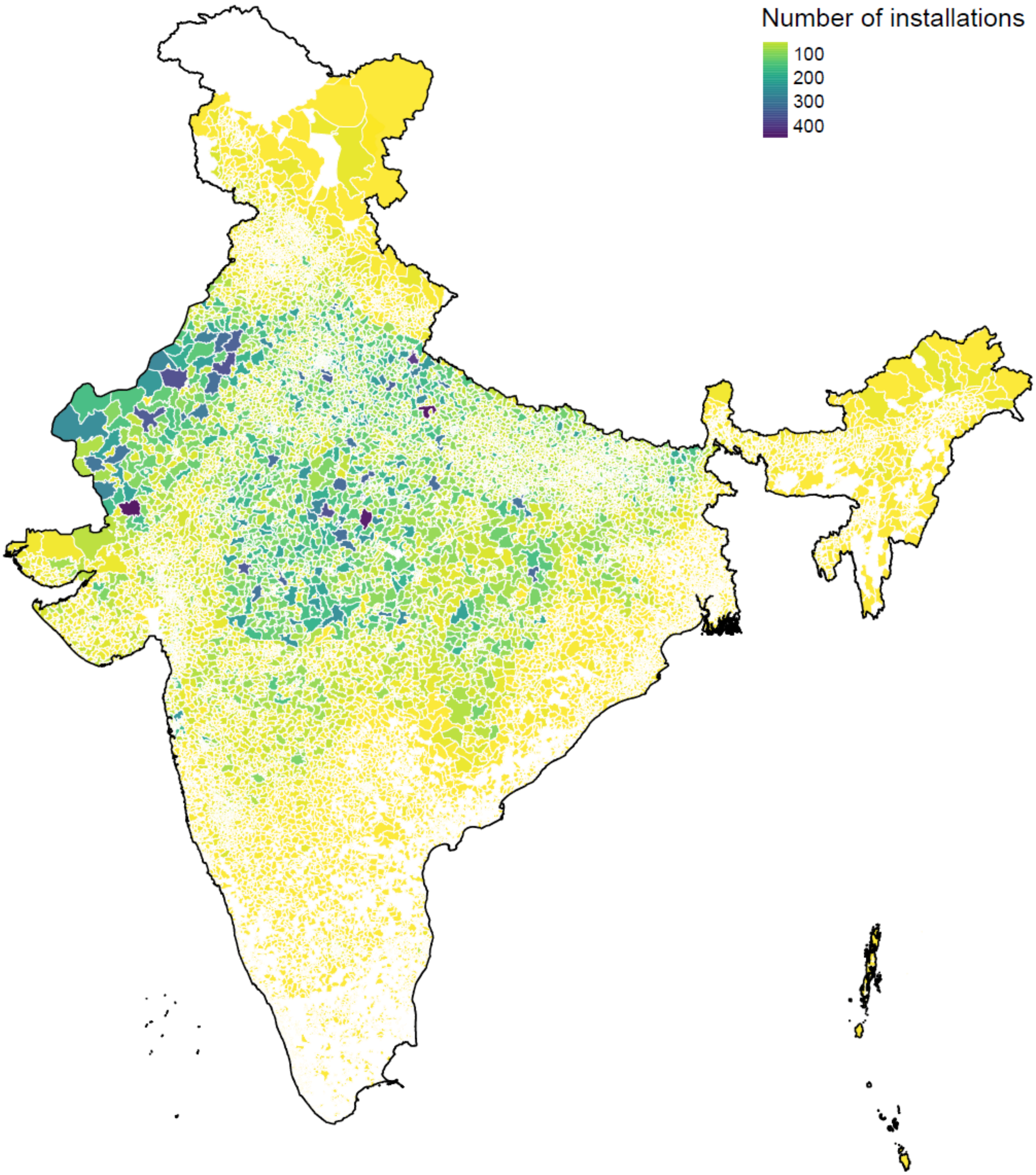
Second, we also obtain a 10% random sample of farmers in the Krishify database along with their detailed search, browsing, like, and comment history. This section discusses the aspects of Krishify database. the random sample also provides other information such as the buckets of land-holdings of the farmers as well as their risk-taking ability. Appendix Figure D.7 presents the distribution of farmers in our sample by their land-holding buckets and their risk-taking ability. Appendix Figure D.8 presents the kernel density of the engagement score that measures the total engagement of the farmer on the platform.

Figure D.5: Monthly Installations of Krishify App



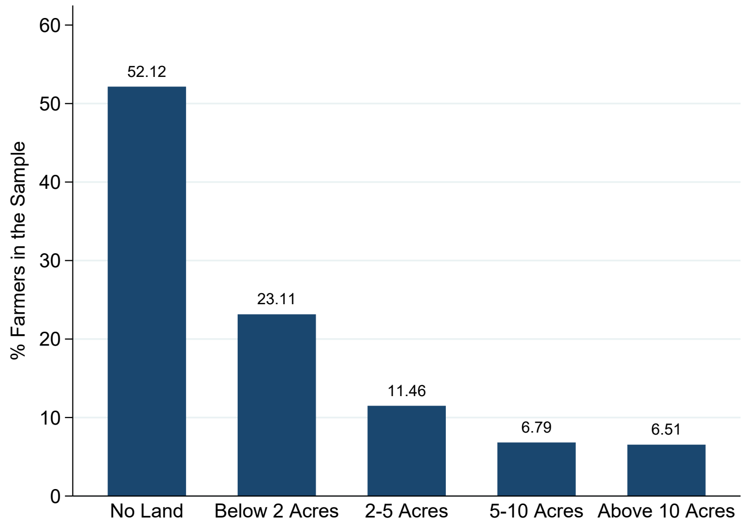
This figure presents the total number of monthly installations of the Krishify app for our sample.

Figure D.6: Geographic distribution of Krishify app installations

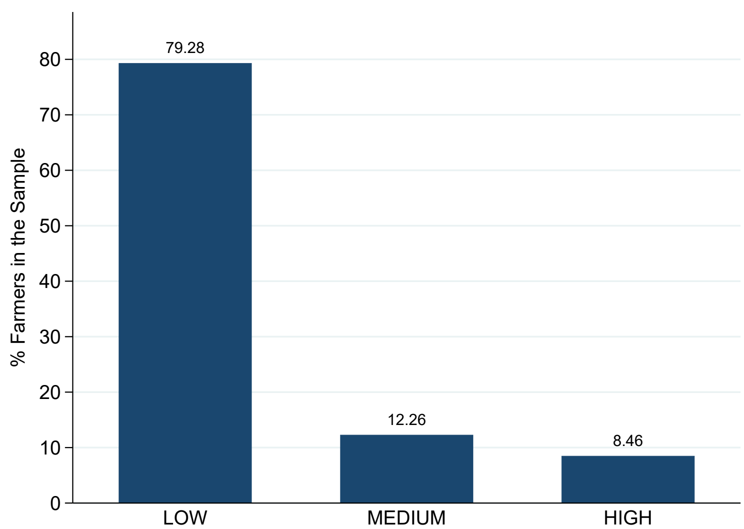


This figure plots a heatmap of the number of total Krishify app installations at the pincode level.

Figure D.7: Description of Farmers in the Krishify Database



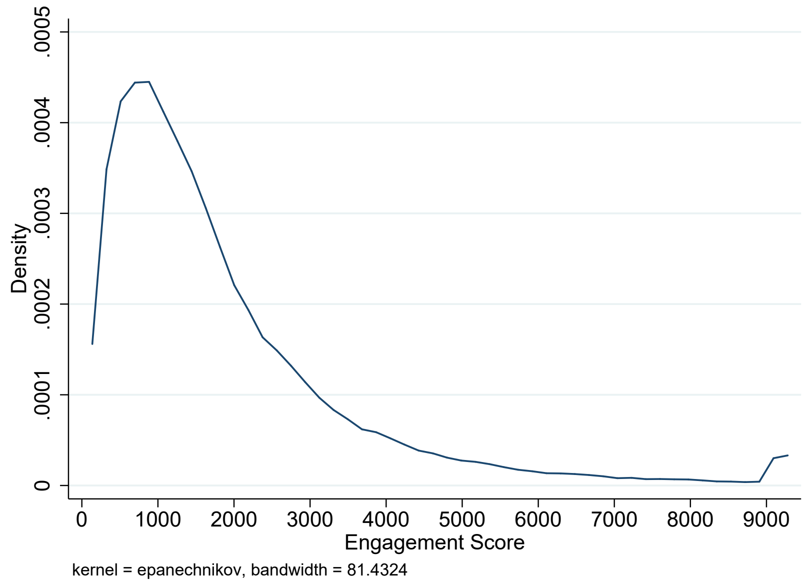
(a) Land Holding



(b) Risk Taking Ability

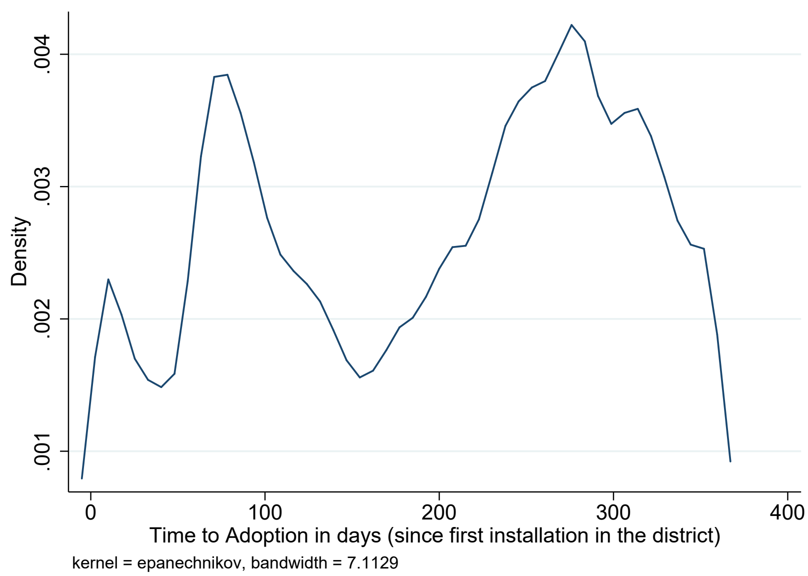
This figure presents the descriptives of the characteristics of farmers in the Krishify database. Figure D.7a presents the distribution of farmers by land-ownership buckets. Figure D.7b presents the distribution of farmers by their risk-taking ability. The sample comprises of a 10% random sample of farmers in the Krishify database.

Figure D.8: Kernel Density of Engagement Score



This figure presents the kernel density of farmer-level engagement score on the Krishify app.

Figure D.9: Kernel Density of Time to Adoption



This figure presents the kernel density of farmer-level time to adoption of the Krishify app. Time to adoption is measured as the duration between the installation date of the app by a farmer and the date of the first app installation in the same district