Watering the Seeds of the Rural Economy: Evidence from Groundwater Irrigation in India

Camille Boudot-Reddy* André Butler†

21st August 2023

Abstract

We study the impact of private investment in groundwater irrigation on the spatial and sectoral distribution of rural economic activity in India. Exploiting a kink in access to groundwater irrigation, generated from an absolute technological constraint on the operational capacity of irrigation pumps with depth of the water table, we find evidence of a significant improvement in agricultural production accompanied with modest consumption gains. Irrigation causes a substantial increase in population density, but has no effect on the employment rate or labour reallocation between sectors of the economy. Furthermore, irrigated agriculture appears to provide additional employment opportunities for waged labour from surrounding non-irrigated villages. Investigating the dynamic effects from adoption indicate important in-migration of labour in the short-run, as well as changes to fertility/mortality in the long to medium-run.

JEL Codes: O12, O33, O53, Q15

*Birkbeck, University of London (c.boudot-reddy@bbk.ac.uk)

†Corresponding Author: University of Cambridge (ajb385@cam.ac.uk)

We thank Liang Bai, Irma Clots-Figueras, Soledad Giardili, Maia Güell, Aprajit Mahajan, Ben Roth, Andreas Steinhauer, and Climent Quintana-Domeque for helpful comments and conversations. We would also like to thank seminar participants at the European Economic Association Annual Consortium (2022), Royal Economic Society Annual Conference (2022), North East Universities Development Consortium Conference (2021), Econometrics Society Delhi Winter Meeting (2021), University of Edinburgh (2021), and the Scottish Economic Society Annual Conference (2021) for their feedback. This paper was previously circulated as "Agricultural Productivity and Local Economic Development: Evidence from Private Investment in Irrigation".

1. Introduction

How a boost to agricultural productivity affects the process of economic growth and development is a long standing question. First chronicled with reference to the Industrial Revolution in England during the 18th century, scholars argued that it was a thriving agricultural sector which enabled subsequent industrialisation (Nurkse, 1953; Robinson, 1954). Building on this evidence, models of structural transformation have shown that a productive agricultural sector can generate demand and hence production in off-farm sectors spurring a movement of labour towards the manufacturing and service industries (Gollin et al., 2002; Ngai and Pissarides, 2007). This view however has been challenged, indicating that in an open economy having a comparative advantage in farming will in fact lead to the pooling-in of labour into the agricultural sector slowing down the development process (Matsuyama, 1992). A resurgence of empirical studies have attempted to shed new light on this debate, demonstrating that the movement of labour between sectors may vary with the technological change (Bustos et al., 2016) and geographic scale (Blakeslee et al., 2023) considered. In this paper, we investigate how groundwater irrigation in India has shaped the rural economy over time.

Irrigation is one of the most conspicuous technologies for stimulating agricultural output. Improved productivity primarily occurs through a direct yield effect, irrigated agriculture is on average at least twice as productive as rainfed (Faurès et al., 2002). Furthermore, the technology has also been found to (1) minimise inter-annual variability by reducing exposure to rainfall shocks (Sarsons, 2015), (2) augment land endowments (Blakeslee et al., 2020), and (3) complement other key inputs such as high yielding varieties (Gollin et al., 2021). In India, advancements in pumping equipment to extract groundwater revolutionised access to irrigation in the early 1970s. In 2013 approximately half of cultivated land across the country was irrigated. Groundwater, accounting for over 70 % of this irrigated land, provides the single largest source of irrigation (Jain et al., 2019). Arguably making this technology one of the most recent salient changes to the agricultural sector.

Groundwater is extracted through tube-wells with an irrigation pump used to move water up the tube to the surface. There are two types of irrigation pump available – centrifugal and submersible. Centrifugal pumps are installed at ground level and generate a pressure differential between the water table and the pumping mechanism. The maximum possible pressure differential at any given altitude is achieved through a perfect vacuum in the pumping mechanism. Under this ideal condition Bernoulli's principle of fluid dynamics dictates that the maximum depth from which water can be extracted by a centrifugal pump is a constant (Faber, 1995). At sea-level this maximum depth is 10.33 meters. Below this threshold, no centrifugal pump will be operational. Extracting water from greater depths requires significantly more expensive submersible pumps which are placed at the bottom of a tube-well and push the water to the surface.

For groundwater depths shallower than the maximum operational threshold, the more cost effective centrifugal pumps are the farmers preferred choice. Hence if all centrifugal pumps were homogenous in their ability to generate a perfect vacuum we would expect to see a jump in access to irrigation at this threshold.[1] Evidence from industry standards however, suggest that centrifugal pumps typically offer a range of efficiencies (Elsey, 2020) such that a jump at any given groundwater depth is unlikely. Instead, there exists a pump efficiency specific threshold such that as we approach the maximum operational depth from shallower levels, a subset of lower-efficiency centrifugal pumps will not function. As we demonstrate empirically, this generates a *kink* in the mapping of centrifugal pump adoption with groundwater depth at the arbitrarily stipulated maximum operational threshold accompanied by an incomplete substitution to the more expensive submersible pumps.

We exploit this quasi-random between village variation in access to groundwater in a *fuzzy* Regression Kink (RK) design. This approach allows us to estimate the causal impact of irrigation on agricultural production and the distribution of economic activity at the local level. Our outcome variables are recorded between 2011 and 2013, by which time half the villages in our sample have had access to irrigation for at least 14 years hence capturing a long to medium-run effect. However leveraging heterogeneity in the timing of tube-well construction also enables us to investigate the dynamic effects from adoption. To test the validity of this empirical strategy formally, we carry out a number of placebo tests to show that groundwater depth does not induce a kink in our outcome variables pre-access to irrigation. Moreover, we demonstrate that there is also no kink with potentially confounding geo-physical covariates such as temperature and rainfall.

In order to leverage spatial variation in groundwater depth at a high-resolution across a large geographic area, we make use of existing and newly-assembled datasets compiled at the village level across the country. Our assignment variable – groundwater depth – is compiled using data published by the Central Ground Water Board (CGWB) which has been monitoring wells four times a year since 1996. We use the geographic positioning system (GPS) locations of these wells to match them to individual villages. Irrigation data, including tube-well construction and ownership of irrigation pumps, is obtained from the Minor Irrigation Censuses; the longest spanning database of information on irrigation infrastructure at the village level. We draw from remote sensing, administrative micro-data, population and economic censuses which cover all households and businesses within the village to measure the local agricultural production, consumption, sectoral labour allocation, and demographics.

We estimate the impact of irrigation as an additional standard deviation unit ($\equiv 103$ $litres/ha/day$) of groundwater on our outcomes of interest. Our results indicate that irrigation significantly improves agricultural production – the monsoon/$Kharif$ agricultural output increases by 7.7%. Additionally, farmers appear to re-optimise their production choices by (1) increasing the share of land cultivated, (2) moving away from drought toler-

[1] Homogeneity in pump efficiency is assumed in the work of Sekhri (2014) when investigating the impact of irrigation on poverty and conflict in India. The author however, takes the maximum pumping depth to be 8 meters based on expert opinion that centrifugal pumps rarely achieve a perfect vacuum.

ant crops, and (3) growing more water intensive crops. Gains in agricultural productivity translate to modest improvements in consumption. We find evidence of an increase in the ownership of household assets, especially solid housing, but no effect on consumption per capita or the poverty rate.

In order to investigate changes in the sectoral distribution of economic activity in the village, we consider employment in the agricultural sector as well as the six largest industries. A productivity boost from access to irrigation does not appear to have transformative effects on the allocation of labour between sectors of the local economy. However when analysing the employment status of residents in the nearest neighbouring village (within 5km of the main sample villages) that has not adopted groundwater irrigation, we find a 17.9% increase in the share of agricultural labourers working full-time. This provides suggestive evidence of a pooling-in of farm labour from less agriculturally productive nearby population centres.

Finally in terms of the village demographics, we find that irrigation causes a large increase in the population density. This appears to be the result of both in-migration, especially by the economically disadvantaged Scheduled Castes, as well as changes in fertility/mortality. Furthermore when considering the dynamic effects of technology adoption, our results are indicative of a significant pooling-in of the Scheduled Caste population in the short-run, increasing by 21.4% in villages having invested in groundwater irrigation post 2000. Conversely, early adopters of the technology that have had access to irrigation for at least 14 years by 2013, demonstrate a significantly higher share of the child population suggesting that changes to fertility/mortality only evolve in the long to medium-run.

Our paper is linked to a resurging literature providing empirical evidence on the effect of productivity shocks in agriculture on the process of economic development. Investigating the role of the Green Revolution on income growth across the developing world, Gollin et al. (2021) finds that the spread of high yielding variety crops significantly increased agricultural productivity, reduced the share of labour in agriculture, thereby initiating the process of industrialisation. Similarly, analysing the effect of an increase in yields from improved fertiliser use in Africa, McArthur and McCord (2017) show that this generated a 14% rise in GDP per capita and led to a 5% decline in the share of agricultural labour over a five year period. In contrast to these studies, our paper exploits high-resolution data with variation at the village level to investigate more localised changes within the rural economy. Our findings indicate that despite villages being at the root of agricultural productivity gains, they do not themselves witness a shift in off-farm opportunities.

At a more micro-level, researchers have attempted to better understand the presence of heterogeneous response to agricultural shocks along different dimensions. In a study exploiting the spread of improved seed varieties in Brazil, Bustos et al. (2016) show that the direction of labour movement between sectors depends on the factor bias of the technological change. The authors find that hybrid maize which enabled a second harvest led to a pooling-in of labour to the agricultural sector, consistent with our findings exploiting irrigation as another form of land-augmenting technology. Using a household-level panel during the peak of the Green Revolution, Foster and Rosenzweig (2004b) suggest important differences on the effect of agricultural gains between landed and landless households. At the village level, we investigate the spatial distribution effects from technology adoption, finding that irrigated villages draw in farm labour from neighbouring non-irrigated villages. Finally our results echo the early work by Foster and Rosenzweig (2004a, 1996) evaluating the impact of the Green Revolution and more recently Blakeslee et al. (2023) and Asher et al. (2022) leveraging variation in access to canal irrigation in India, which all document a lack of village level off-farm growth following production gains in the agricultural sector. We add to this literature in two ways. First, most studies have focused on technological change dating back to the 1960s in the case of the Green Revolution and even earlier for canals. In contrast our paper studies a much more recent period of agricultural change in response to groundwater irrigation. Secondly, we exploit heterogeneity in the timing of access to this technology to investigate the dynamic effects from adoption.

There also exists a number of studies which have evaluated transient agricultural shocks due to climate variability. For instance, consistent with a theory of demand driven structural change, Emerick (2018) finds that positive rainfall shocks in India which boost agricultural production also lead to an expansion of the non-tradable sector. In response to negative weather shocks in China, Minale (2018) reports in a significant proportion of the agricultural labour force migrating to urban centres. In comparison to this literature, our paper studies how the local economy adjusts to a persistent shift in agricultural productivity.

Our work also adds to a strand of causally interpretable evidence on the impact of irrigation. The scarcity of such research is due in large part to the empirical challenges involved in establishing reliable estimates. In a seminal paper, Duflo and Pande (2007) analyse the distributional effects of dams in India finding that beneficiaries living downstream increased their agricultural productivity and experienced lower levels of poverty. Access to irrigation has also been documented to reduce conflict (Sekhri, 2014), and encourage farmers to move away from cultivating drought tolerant crops which increases vulnerability to climate shocks in the long-run (Hornbeck and Keskin, 2014). Using randomly located geological formations that store pockets of water in the bedrock, Blakeslee et al. (2020) explore farmer adaptations to the drying up of groundwater for irrigation. The authors find that while there is an immediate consequential decline in farm income, households appear to successfully offset these losses by reallocating labour to off-farm employment.

The rest of the paper is structured as follows. Section 2 provides insight on the use of irrigation in India over time and describes the different technologies available to farmers for groundwater extraction. Our data sources are explained in Section 3, and the empirical strategy including graphical evidence is presented in Section 4. Section 5 contains results on the impact of irrigation on the rural economy. Finally, Section 6 concludes.

2. Background

In the 1950s, following independence, India invested extensively on public provision of irrigation infrastructure making canals the dominant source of water for agricultural purposes (Jain et al., 2019). However over the years, mininal maintenance of the infrastructure resulted in water supply from these canal networks becoming increasingly unreliable (Mukherji, 2016). At the same time, technological advancements in pumping equipment accompanied by government energy subsidies to operate these pumps made extracting groundwater an affordable and appealing option (Shah et al., 2012). Hence as of the early 1970s, groundwater overtook canals as the largest source of irrigation. The following decades witnessed a groundwater revolution – by 2013 groundwater accounted for 70% of the country's irrigated area while the share irrigated by canals had declined to 20% (Jain et al., 2019).

Figure 1 shows this gradual evolution in groundwater extraction over time among the sample of villages used in this study. The share of villages with tube-wells increased five-fold between 1986 to 2013. This expansion is also reflected on the intensive margin of technology adoption over this period – on average, the number of tube-wells used to extract groundwater for irrigation increased from 3 to 52 per village. This implies that by 2013, which is when our primary outcome variables are recorded, half of the villages will have had tube-wells for at least 14 years. Hence our study captures the medium to long-run impact of private investment in groundwater irrigation, conceivably one of the most salient recent technological innovation aimed at boosting agricultural productivity (Mukherji, 2016).

Groundwater is extracted through tube-wells. A tube-well consists of a bore hole which is drilled into the ground so as to tap groundwater from porous zones in the aquifer. An irrigation pump is then used to move the water up the tube to the surface. There are two main types of irrigation pump available – centrifugal and submersible. The choice of which pumping technology is most suitable for extracting groundwater depends on the depth of the water table in that location.

Centrifugal pumps are installed at ground level and create a vacuum with water moving up the tube from an area of high pressure at the bottom of the tube-well, to an area of low pressure in the pumping mechanism (Figure 2). The extraction of water from a tube-well using a centrifugal pump can be described by Bernoulli's principle of fluid dynamics (Faber, 1995) (Equation 1):

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2 \tag{1}$$

where the variables P_i, v_i, and h_i refer respectively to the pressure ($kg/m/s^2$), velocity (m/s), and height (m), between the pump ($i = 2$) and the water table ($i = 1$). The constants, ρ and g are the density of water (997 kg/m^3) and gravitational force (9.81 m/s^2) respectively. Assuming constant flow velocity we can rewrite Equation 1 in the following form:

$$h_2 - h_1 = \frac{P_1 - P_2}{\rho g} \tag{2}$$

As can be interpreted from Equation 2, the maximum possible pressure differential is achieved through a perfect vacuum (P_2=0 $kg/m/s^2$) in the pumping mechanism. Under this ideal condition and atmospheric pressure at sea-level (P_1=101,325 $kg/m/s^2$) the maximum depth from which water can be extracted – that is, the difference between h_2 and h_1 – is 10.33 meters. This represents the maximum theoretical threshold achievable by a centrifugal pump.

Realistically however, it is unlikely that all centrifugal pumps are able to create a perfect vacuum. Industry standards suggest that centrifugal pumps more typically offer efficiencies ranging from 55 to 93 percent (Elsey, 2020).[2] This will reduce the depth from which a centrifugal pump can extract groundwater. In Figure A1, we show that at sea-level the depth from which a centrifugal pump can extract water falls from 10.33 to 5.18 meters as pump efficiency falls to half its maximum potential. Therefore, we propose an efficiency specific threshold below which a centrifugal pump can no longer be used to access groundwater for irrigation.

In a scenario where a centrifugal pump can no longer operate, submersible pumps can provide an alternative technology for water extraction. Submersible pumps are placed at the bottom of the tube-well and push the water to the surface. Consequently, provided it has sufficient horsepower, a submersible pump could extract water from any depth.

Given its additional functionality, a submersible pump is significantly more expensive than a centrifugal pump. Based on an online search among India's top five irrigation pump manufacturers, we found that the starting price of centrifugal pumps was less than half that of submersible pumps.[3] The lowest priced centrifugal was 3,000 Rupees (30 GBP) compared to 7,500 Rupees (75 GBP) for the lowest priced submersible pump.[4] To put these costs into context, the mean annual per capita consumption in our sample of villages is approximately 18,000 Rupees (GBP 180).

In Appendix B we provide a simple decision making framework for the adoption of these different irrigation pumping technologies available to farmers and demonstrate that

[2] We verified this on the site of numerous irrigation pump suppliers and manufacturers. The information indicated that centrifugal pumps could achieve up to 90% efficiency, with most pumps ranging from 50 to 80%. See for instance: `https://www.tapflopumps.co.uk`, `https://www.rotechpumps.com`, and `https://www.inoxmim.com`.

[3] We sourced this information from providers, including: `https://www.moglix.com` and `https://www.indiamart.com`.

[4] As an additional comparison we verified prices for the top three selling centrifugal and submersible pumps. While centrifugal pumps ranged from 4,500 to 5,700 Rupees (45-57 GBP), the top three selling submersible pumps were priced between 10,000 to 12,000 Rupees (100-120 GBP). Similarly when comparing prices for pumps with the same features (e.g. horse power), submersible pumps where also twice the price of centrifugal pumps.

it is the subset of farmers that can afford a centrifugal pump but not a submersible that generate a decline in overall centrifugal pump adoption culminating in zero take-up at the maximum theoretical threshold. We empirically demonstrate the presence and validity of this relationship in Section 4.

3. Data

For the purpose of this study we link observational groundwater data from wells in 2013 with multiple external contemporaneous datasets describing irrigation practices and the rural economy to obtain a village level cross-section. Importantly for our empirical approach, this enables us to leverage spatial variation in irrigation at a high-resolution over a large geographical area.

3.1 Groundwater

Data on our assignment variable - groundwater depth - come from the official website of the Central Ground Water Board (CGWB).[5] Since 1996, the CGWB has kept digitised records from groundwater monitoring wells evenly spread across the entire country. In 2013, the CGWB had a total of 17,116 monitoring wells covering 511 districts across 21 States.[6] Wells are identified by Global Positioning System (GPS) coordinates and are monitored four times in the year – pre-monsoon, mid-monsoon, pre-winter, and post-winter[7] – so as to capture both seasonal and inter-annual variation.

We construct our assignment variable as the maximum groundwater depth recorded at any point over a three year period (2010-2013).[8] Taking a three year horizon enables us to account for some of the temporal fluctuation which may affect groundwater depth. There are concerns that groundwater irrigation in India is leading to a depletion of its aquifers and the water table falling over time (Famiglietti, 2014). However as Figure 1 demonstrates, between 1996 to 2013, the annual average maximum groundwater depth among our sample of villages was stable around 7.5 metres. If however we consider the complete sample of villages, that is including those outside our bandwidth, we do observe a drop in the average maximum groundwater depth from 8 to 10.5 metres between 1996 to 2013.

Combining village boundary shapefiles offered by the Socioeconomic Data and Applications Center (SEDAC) of NASA,[9] along with the GPS coordinates of wells, we create a village level match. Specifically, we attribute the measure of our assignment variable to a

[5] Data can be downloaded in excel format from: http://cgwb.gov.in.

[6] In 2011, India had a total of 640 districts across 28 States. The CGWB therefore provided coverage for over 80% of India's districts at the time.

[7] With some regional variation, the monsoon/*Kharif* season is from June to October and the winter/*Rabi* season is from November to March.

[8] Of the total groundwater monitoring wells sampled by the CGWB not all are monitored every year. As a result, our assignment variable can only be calculated for a subset of 8,549 wells.

[9] Shapefiles mapping the whole of India are available at: https://sedac.ciesin.columbia.edu/data/set/india-india-village-level-geospatial-socio-econ-1991-2001.

village if the well falls within the village boundary. If more then one well was matched to the same village, an average of our assignment variable was taken. Figure 3 presents a map of our final sample of matched wells across the country as well as whether these fall below or above the maximum theoretical threshold of a centrifugal pump. This threshold is calculated based on Bernoulli's principle of fluid dynamics described in Equation 2, assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude.[10] As Figure 3 plainly demonstrates, our data on groundwater depth provide the basis of our empirical approach – evenly distributed high-resolution spatial variation across a large geographic coverage.

3.2 Irrigation

We compile data on irrigation practices from the Minor Irrigation (MI) Censuses conducted every 7 years since 1986 for the planning and management of water resources in the agricultural sector.[11] These Censuses provide a countrywide database of groundwater and surface water infrastructure that have a culturable command area of less than 2,000 hectares – known as minor irrigation schemes.[12]

Specific to the needs of our study, the Fifth MI Census (2013) has data on ownership of different pump types, including submersible and centrifugal. Importantly, there also exists information on pump capacity (horse power) and usage (pumping hours) which we leverage to calculate water input in litres following a standard engineering formula (Manring, 2013) (for a complete discussion on the construction of this variable see Appendix C). This measure enables us to capture the intensive margin of access to irrigation. Ryan and Sudarshan (2022) also make use of this water use intensity variable in their recent work evaluating the effect of groundwater rationing in Rajasthan.

Furthermore, we make use of the historical record of tube-well construction from all the MI Censuses to map the evolution of technology adoption over time. As demonstrated in Figure 1, by 2000 approximately half of the villages in our sample have had access to irrigation for at least 14 years. We exploit this heterogeneity in the timing of adoption to investigate the dynamic effects from access to irrigation.

[10] Data on altitude was extracted from raster files obtained from the ALOS Global Digital Surface Model for the whole of India. A barometric formula was used to calculate atmospheric pressure at varying altitude. Specifically, we estimate $P = P_b.exp[\frac{-gM(h-h_b)}{RT_b}]$, where P refers to pressure, g is the gravitational force, M is the molar mass of the Earth's air, h is height, R is the universal gas content, and T is temperature. Note that though the base values for P_b, h_b and T_b naturally evolve with altitude, these are in fact constant for the range of altitude found in our sample.

[11] Village level data from the MI Censuses are publicly available in excel format on the Government of India open data platform at: `http://data.gov.in`. Background information on each Census (e.g. questionnaires and instruction manuals on data collection) as well as official reports and aggregated statistical tables can be found on the official website of the MI Census at: `http://micensus.gov.in`.

[12] In contrast, medium and large irrigation schemes have a culturable command area of 2,000-10,000 ha and above 10,000 ha respectively. These include dam and canal irrigation infrastructure.

3.3 Agriculture

Data on agricultural production based on direct field measurements is not available at the village level. We therefore rely on measures of vegetation cover calculated from satellite images as a proxy for agricultural production at the village. Specifically, we use data from the Enhanced Vegetation Index (EVI) estimated from images taken by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard NASA's Terra satellite. This data was used by Asher and Novosad (2020) to evaluate the impact of India's national rural road expansion programme and was made available by the authors as part of their replication dataset.[13] The authors extracted information on EVI from gridded datasets across India for nine 16-day periods from June to October – covering the monsoon/$Kharif$ growing season – and similarly from November to March – covering the winter/*Rabi* growing season – over a fourteen year period (2000-2014). This data was then aggregated to our unit of analysis using village boundary shapefiles. We use the maximum EVI value (log transformed for ease of interpretation) in each agricultural season of 2013 as our preferred outcome variable.[14] Appendix C includes further information on how the indices and proxies are constructed, a discussion of the literature on using remote sensing imagery to predict crop production, as well as results from validation tests showing the correlation between the indices and district level estimates of agricultural production.

In this study we are not only interested in capturing changes to agricultural production in response to irrigation, but importantly changes in agricultural production choices. We therefore leverage data from multiple external sources in order to obtain a range of village level indicators on input use and crop choice. The Village Directory, administered as part of the 2011 Population Census, keeps records of the three principle crops grown in each village.[15] We use this information to create three binary measures of crop choice. Specifically, does a village grow water intensive, drought tolerant, and cash crops.[16] In terms of agricultural inputs, we also draw upon the 2011 Village Directory for our measure of land area cultivated. Finally, we compile data on two indicators of technology adoption – water-saving technology (drip and sprinklers) is obtained from the Fifth MI Census (2013) and mechanised farm equipment collected as part of the Socio Economic Caste Census

[13]The paper by Asher and Novosad (2020) and its associated dataset is available at: `https://www.aeaweb.org/articles?id=10.1257/aer.20180268`

[14]We demonstrate robustness of our results for varying specifications of this variable, including using maximum EVI in level form as well as measuring the maximum as an average over three and five years. Additionally, we report results on another proxy used in crop-mapping studies which measures the difference between early season EVI and the maximum value.

[15]Data from the 2011 Population Census Village Directory can be downloaded from: `https://censusindia.gov.in/2011census/censusdata2k11.aspx`.

[16]Based on classification by the International Crops Research Institute for Semi-Arid Tropics, water intensive crops include sugarcane, cotton, and rice, while drought tolerant crops include millet, sorghum, maize, pigeon pea, and groundnut. Cash crops include sugarcane, oilseed, cotton, and tobacco. These crops cannot be directly used for household consumption as they require post-harvest processing, but are generally considered to be more profitable.

(SECC) of India in 2012.[17]

3.4 Consumption

We consider a range of income and expenditure indicators. These are all obtained from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG, Version 1.5).[18] Using household micro-data collected by the India Human Development Survey-II in 2012, the SHRUG provides predicted variables for consumption per capita and the poverty rate based on a range of asset and income variables equivalent to those recorded in the SECC (2012). Following the methodology suggested by Elbers et al. (2003) and explained in detail in Appendix C, we leverage these predictions to impute consumption and poverty using the SECC asset and income data so as to generate village level statistics. Additionally, we consider an index of asset ownership as recorded by the SECC (2012), as well as each individual major asset independently. Finally, we calculate a measure of average night light (log transformed for ease of interpretation) per village in 2013. Night light, captured by satellites as the pixel luminosity in a geographic polygon, is widely used as a proxy for economic activity when direct measures are otherwise unavailable (Henderson et al., 2011). Appendix C provides for a detailed discussion of the construction of this variable, as well as evidence from the literature on its correlation to other economic indicators.

3.5 Labour

We draw on the 2011 Population Census for information on labour allocation of village residents. Specifically, we obtain data on total employment, as well as for two occupational categories of employment in the agricultural sector – cultivators and labourers. Cultivators are those that cultivate their own land, while labourers work for a daily wage. Data on these categories is available disaggregated by gender, enabling us to test for shifts in labour allocation for men and women separately. Furthermore, the data can also be disaggregated by time spent employed. The Census of India considers two types of workers – main/full-time workers are defined as those that are economically active in an employment category for more than 6 months of the year, while marginal/part-time workers are active for less than 6 months.

So as to obtain information on employment in village industries other than the agricultural sector, we make use of data from the Sixth Economic Census conducted in 2013.[19]

[17] The Government of India regularly conducts SECC surveys at the individual and household level so as to determine eligibility into social programmes. Village level aggregates of this survey, including household assets, are made available online as part of the work of Asher and Novosad (2020) evaluating India's national rural road construction programme. As mentioned previously, this paper and its associated dataset is available at: `https://www.aeaweb.org/articles?id=10.1257/aer.20180268`

[18] For detailed information on the SHRUG, please refer to Asher et al. (2021). The dataset, including codebooks and references, can be found at: `http://www.devdatalab.org/shrug`

[19] This data is available on the National Data Archive site: `http://microdata.gov.in/nada43/index.`

The Economic Census is the only complete enumeration of all economic establishments in India, formal and informal, with no restrictions on size or location.[20] Detailed records are kept on employment and business characteristics including industry classification. We concentrate our analysis on employment in the following village industries: agro-processing (this excludes crop production), livestock, education, manufacturing, services, and forestry. Among our sample, these six industries account for over 85% of employment. Unfortunately, the Economic Census does not ask any questions on wages, inputs, or outputs, hence we cannot investigate shifts in the production or profitability of these industries.

3.6 Demographics

Using the Population Census of 2011, we calculate population density of a village as the ratio of total population to village area. Additionally, we disaggregate this measure by gender and age. The adult population includes all those aged 18 years and over,[21] while the child population is between 0 to 6 years. In an attempt to identify the presence of in-migration versus shifts in fertility and/or mortality, we calculate the share of the total population by age and gender. Finally, we consider the share of the Scheduled Caste population. Members of these castes are among the most disadvantaged and form a significant share of migration within the country.[22]

3.7 Covariates

In a placebo test, we consider four covariates which capture natural geo-physical features of the village: temperature, rainfall, distance to nearest river, and whether the village is in the command of a canal network (which depends on the local topography). The Climate Hazards Centre makes publicly available quasi-global high-resolution gridded datasets on temperature (CHIRTS) and rainfall (CHIRPS) (see Funk et al. (2014) and Funk et al. (2019) for details on how these datasets are compiled using both satellite measures and on-site station records).[23] From these files, we extract information on the maximum monthly temperature as well as annual rainfall and match these to villages using our village boundary shapefiles. In order to account for temporal weather fluctuations we compute an average of these measures over a three-year period (2010-2013). Distance to the nearest

`php/catalog/47`.

[20] An establishment refers to any unit where an economic activity is carried out; with the exception of those engaged in crop production, defence, and government administration.

[21] At the village level the Population Census reports data on the total population as well as the population aged 0-6 years. We estimate the 0-18 years population by multiplying the 0-6 population by 18/7. We then estimate the adult population as the total minus the 0-17 years population.

[22] According to the 2011 Population Census, 16% of intra-State migrants belonged to the Scheduled Castes.

[23] Temperature data and related information is available at: https://www.chc.ucsb.edu/data/chirtsmonthly, while precipitation data and its related information is available at: https://chc.ucsb.edu/data/chirps.

river is obtained from the 2011 Village Directory. Whether or not the village is in the command area of a medium to large irrigation scheme, including dams and canal networks, is taken from the Fifth MI Census (2013).

Finally, we show a balance test of our key outcome variables pre-access to irrigation. For this, we rely on a sub-sample of villages with tube-wells built solely after 2000 and data on our outcomes prior to this date. Irrigation is measured using the Third MI Census conducted in 2000, village demographic indicators are obtained from the 2001 Population Census of India, and proxies of consumption come from the 2002 Below Poverty Line Census. Additionally, we make use of the previously described satellite imagery proxies for agricultural production and night light dating back to 2000.

Each part of the dataset outlined above is matched at the village level. For variables coming from the Population Census and Economic Census these are directly matched to the SHRUG Dataset of India (Version 1.5) using Census village identifiers. As explained previously, groundwater and weather data is linked using village shapefiles. Finally, for the irrigation data we use a combination of Python and Stata code for fuzzy matching on names adapted to the local Indian languages.[24] Our final sample of villages are those that have: (i) non-missing information across all our variables, (ii) tube-wells built by 2013, and (iii) groundwater depth within the bandwidth of 7 metres from the maximum theoretical depth threshold of a perfectly efficient centrifugal pump.[25] Table 1 provides descriptive statistics of all key variables on the final sample size of 3,327 villages across 423 districts in 19 States of India. These statistics suggest that in an average village, approximately 76% of the agricultural land is irrigated by tube-wells. Agriculture is the largest employer with approximately 13% of the workforce engaged as cultivators and 18% as manual labourers. The village population encompasses just over 4000 people, with 20% belonging to the Scheduled Castes and 28% recorded as being below the poverty line.

4. Empirical Approach

In this paper, we are interested in capturing the effects of irrigation on agricultural production and the distribution of local economic activity. Irrigation practices however, are likely to be endogenous. For instance, we might expect that villages with better access to markets are more likely to adopt tube-wells. Any naive correlation estimates between groundwater extraction for irrigation and economic outcomes will in such a case be biased; partially attributing the effect of irrigation to markets rather than the technology itself.

[24] We use the *Masala Merge* algorithm developed by Paul Novosad which modifies the Levenshtein edit distance to lower the cost of certain substitutions that are common to Hindi. The code and information on this algorithm is available on the authors' website: `https://www.dartmouth.edu/~novosad/code.html`. This method resulted in a matching success of approximately 80%.

[25] The maximum theoretical threshold of a centrifugal pump is calculated based on Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude

In order to identify exogenous variation in irrigation, we exploit the laws of physics which dictate that there exists an arbitrary maximum groundwater depth from which water can be extracted by a centrifugal pump.

Previous work by Sekhri (2014) evaluating the effect of access to water on poverty and conflict in rural India also used the physical constraint on the operational capacity of centrifugal pumps with groundwater depth as a source of exogenous variation. The author adopts a *fuzzy* Regression Discontinuity (RD) Design at a threshold of 8 meters based on expert opinion that achieving a perfect vacuum in the pumping mechanism is in practice unlikely. However reports from industry standards suggest that centrifugal pumps in fact typically offer efficiencies ranging from 55 to 93 percent (Elsey, 2020). Hence a *jump* in access to groundwater, whether at 8 or 10.33 meters or anywhere in between, is unlikely. In contrast we propose that pumps are drawn from a distribution of efficiencies leading to a gradual decline in adoption of the technology culminating in zero take-up at the maximum theoretical threshold, hence generating a *kink* in access to groundwater at that point.

In this section we outline our proposed empirical approach – *fuzzy* Regression Kink (RK) Design.[26] Furthermore, we present graphical evidence and estimation results corroborating the validity of this method.

4.1 Regression Kink Design

Centrifugal pumps provide the most affordable technology to privately access irrigation. However as described in Section 2, there exists a maximum theoretical threshold below which a centrifugal pump can no longer operate. Furthermore, the functionality of a pump is limited by its efficiency. Hence as we approach the maximum theoretical threshold from shallower depths, a subset of the lower efficiency pumps will not be operational. This leads to a gradual decline in the use of centrifugal pumps for irrigation, with zero take-up of the technology at the maximum theoretical threshold.[27] In this context the change in slope of the assignment function, which maps the relationship between groundwater depth and irrigation, at the kink point is unknown and must be estimated based on observed data. Accordingly, we employ a *fuzzy* RK design[28] (Card et al., 2015b) wherein the assignment function is specified as:

[26] There has been increasing interest in adopting RK designs in the applied economics literature. The most common application so far has been the use of kinks in unemployment benefit schedules to capture the effect of these on labour market outcomes (Card et al., 2015a; Landais, 2015). A small but growing literature has also used this method to evaluate a range of topics including, but not limited to, the effect of coalition governments on fiscal policies (Garmann, 2014), financial aid on educational outcomes (Nielsen et al., 2010), and demand for prescription drugs (Simonsen et al., 2016).

[27] See Appendix B for a formal presentation of a decision-making framework explaining the adoption of different pumping technologies by farmers.

[28] The difference between a *sharp* and *fuzzy* RK design is that the *fuzzy* RK design estimand replaces the known change in slope of the assignment rule at the kink point with an estimate based on the observed data.

$$I_{vds} = \delta_0 + \delta_1(w - k) + \delta_2(w - k) \cdot D_{vds} + \sigma X_{vds} + \eta_s + \varepsilon_{vds} \tag{3}$$

I_{vds} is groundwater extracted for irrigation in village v, district d, and state s. w is the groundwater depth. k is the kink point, calculated based on Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude. D_{vds} is a binary indicator which takes the value one if village v has a groundwater depth w below the kink point k; that is $w > k$. We expect a kink in the deterministic relationship between our treatment variable, irrigation, and our assignment variable, groundwater depth, at k. It follows that if irrigation exerts a causal effect on our outcome of interest we should then also expect to see an induced kink in the relationship between the outcome and our assignment variable at k. This outcome function is estimated as:

$$Y_{vds} = \gamma_0 + \gamma_1(w - k) + \gamma_2(w - k) \cdot D_{vds} + \nu X_{vds} + \mu_s + \upsilon_{vds} \tag{4}$$

Y_{vds} is the outcome of interest. The causal impact can then be calculated as the ratio of the coefficients – $\beta = \gamma_2/\delta_2$ – and interpreted as the average treatment effect on the treated. Standard errors for β are recovered using the Delta method. All our regressions use a bandwidth of 7 metres from the kink point and a linear functional form. However we demonstrate that our results are robust to a range of bandwidth down to 3 metres and compare our findings when using a quadratic and cubic function.

Control variables and fixed effects are not necessary for identification in an RK design, but do improve the efficiency of the estimation (Calonico et al., 2014; Imbens and Lemieux, 2008). We therefore include a vector of village geo-physical covariates, X_{vds}, which include temperature, rainfall, distance to river, and whether the village is in the command area of a canal as controls in our specification.[29] Furthermore, we also include state fixed effects, η_s and μ_s in Equation 3 and Equation 4 respectively. We show in a robustness test that excluding these controls does not affect our results.

4.2 Impact of Groundwater Depth on Irrigation

Identification in a *fuzzy* RK design requires three key assumptions (Card et al., 2015b): (1) the conditional density of the assignment variable, given the unobserved error in the outcome, is continuously differentiable at the kink point, (2) there is no jump in the direct marginal effect of the assignment variable on the outcome of interest at the kink point,[30] and (3) covariates are continuously differentiable at the kink point.

In response to the first assumption we plot the probability density function of the

[29] These covariates are shown not to exhibit a discontinuity in the first derivative of the assignment function at the kink point.

[30] As explained by Card et al. (2015b), this condition is what differentiates an RK to an RD design. In absence of this condition, wherein there exists a jump rather than a kink, an RD design would be used.

assignment variable, groundwater depth, to check for manipulation of ones' position at the kink point. Firstly we note that as shown in Panel A of Figure 4 the exact location of the kink point is village specific as it is adjusted for the local altitude. Panel B of Figure 4 shows the number of observations in each bin for groundwater depth normalised at the kink point. The evolution of the distribution of our assignment variable shows no signs of discontinuity around this point. This is further supported by the McCrary test, commonly used in the RD literature, which estimates the log change in height between bins at the kink point. Results from this test (displayed directly on the graph) confirm that we cannot detect a significant discontinuity at that point.

The second assumption validates the treatment effect. Corroborating the known technological constraint and the effect of efficiency in limiting the operation of centrifugal pumps with groundwater depth, Panel A of Figure 5 demonstrates a clear kink in the slope of the relationship between centrifugal pump adoption and groundwater depth normalised at the kink point. Specifically, we find a decline in the adoption of centrifugal pumps as groundwater depth increases followed by a sharp visible switch to a constant near zero adoption at the kink point ($w > k$). As expected, the price differential of submersible pumps limits the substitution to this alternative technology (see Panel B of Figure 5). As such, the amount of water extracted for irrigation closely follows the same change in slope as centrifugal pump adoption with groundwater depth (see Panel C and D of Figure 5).[31] This graphical evidence is further substantiated in Table 2 which presents our results on the assignment function. These indicate a statistically significant positive change in the slope of centrifugal pump adoption (Column 1) with groundwater depth at the kink point, and similarly in the case of irrigation (Columns 3 to 5).

Finally, the third assumption attempts to address the concern that there may be village characteristics which are correlated to the treatment status. We first demonstrate that for a range of covariates capturing local geo-physical factors, these do not exhibit a kink in their relationship with groundwater depth. For instance, as shown in Panel A and B of Figure 6 (with formal RK estimates reported in Columns 6 and 7 of Table 2) there is clearly no kink in the relationship between groundwater depth with either temperature or rainfall. Secondly, we test for the balance of our key outcome variables – irrigation, agricultural production, poverty, and population – prior to having access to groundwater irrigation. For this, we observe the distribution of these variables between 2000 to 2002 depending on the source of data among a sub-sample of villages that built tube-wells solely after 2000. Table 3 presents the mean for this sub-sample of villages (Column 1), as well as disaggregated for villages just below (Column 2) and those just above (Column 3) the kink point. While there are average differences between villages (Columns 4 and 5), we find no statistically significant change in slope at the kink point when using the *fuzzy* RK

[31] The cost of operating a submersible pump is significantly more expensive than a centrifugal. Among our sample of villages, we find that average number of pumping hours for centrifugal pumps during the *rabi* season is more than double that of submersible pumps. This may explain the low correlation between water extraction and pump adoption for deeper depths.

specification (Columns 6 and 7).

Robustness: In support of our empirical strategy, we conduct various robustness tests reported in Appendix A. First, presented in Table A1, we demonstrate that our results are consistent when the assignment variable is calculated as the maximum groundwater depth observed over a one, three, or five year time horizon. Second, reported in Table A2 we analyse the sensitivity of our results to the choice of polynomial order. The AIC is very similar across all specifications, but we find that the standard errors increase substantially with higher order polynomials (this trend has been noted in other RK studies; see Landais (2015)). Third, we explore the sensitivity of the deterministic relationship between irrigation and depth of the water table to the choice of bandwidth level. As shown in Figure A2, our results are consistent across bandwidth down to 4 meters either side of the kink point.

5. Results

In this section we report and discuss our results on the impact of an agricultural productivity shift from access to irrigation on the sectoral distribution of rural economic activity. For each outcome variable we report the beta estimate (with the heteroskedasticity robust standard errors in brackets) corresponding to the ratio of the coefficients capturing the change in slope of the outcome (Equation 4) and the assignment function (Equation 3) at the kink point (explained in Section 4). Our treatment variable, irrigation, is measured as water extracted in $litres/ha/day$ and standardised such that all results can be interpreted as the effect of a one standard deviation ($\equiv 103$ $litres/ha/day$) increase in irrigation.

5.1 Agriculture

Before all else, we evaluate the impact of irrigation on agricultural production. To do so, we leverage the maximum Enhanced Vegetation Index (EVI) value calculated from satellite imagery as a proxy for agricultural yields in both the monsoon/$Kharif$ and winter/*Rabi* season of 2013. Irrigation appears to have a positive effect in fostering agricultural production, especially during the monsoon/$Kharif$ season. We estimate that a one standard deviation increase in irrigation significantly boosts agricultural production by 7.7% during the monsoon months, reported in Column 1 of Table 4. This result is robust to variations in the calculation of the EVI proxy (reported in Table A3).[32] Additionally, we also present graphical evidence of this effect in Panel A of Figure 7. Similar to that detected for centrifugal pump adoption and irrigation, there is a sharp decline in monsoon/$Kharif$

[32] Table A3 reports results when using variations in the computation of the EVI proxy. Specifically, we show that the estimates are robust to (1) using the level form of the EVI maximum value, (2) calculating the maximum over varying time horizons in order to account for temporal fluctuations and potential outlying years, and (3) using the difference between early season and the maximum value (a proxy also commonly used in crop-mapping studies with satellite imagery).

agricultural production with groundwater depth up until the kink point and levelling off at greater depths.

Having established the effect of irrigation on agricultural production, we go on to analyse the pathways through which these effects may operate over and above the direct yield impact. Improvements in agricultural output could happen through two main channels: (1) conditional on higher production translating to higher profits farmers may increase investment in other inputs, and/or (2) farmers may re-optimise their production strategy in response to a reduced exposure to climate risk.

In response to the first channel we investigate investments in a range of inputs including land, water-saving technology, and mechanised equipment, reported in Columns 3 to 5 of Table 4 respectively. Irrigation significantly increases the share of village area used for cultivation. A one standard deviation increase in irrigation leads to an 18.7% rise in the proportion of village land being cultivated (Panel B of Figure 7 provides corresponding graphical evidence). We may expect that alongside this increased land use, there is also a shift in the size of landholdings. For instance, larger farmers may be more likely to invest in irrigation and buy out smaller less productive non-irrigated farms. To test for this we analyse the impact of irrigation on the share of households in four different landholding categories – landless, 0-2, 2-4 and above 4 acres, reported in Table A4. We find no evidence of a shift in landholding size, likely a reflection of how thin land markets are in rural India (Mearns, 1999). We also do not detect any shifts in the ownership of mechanised equipment, and a marginally significant decline in the use of water-saving technology. This latter result confirms the growing concerns that private investment in tube-wells without any regulations on treating groundwater as a common resource may be a leading factor in water mismanagement (Dubash, 2007).

With respect to the second channel we analyse shifts in the most common crops grown in the village, presented in Columns 6 to 8 of Table 4. We consider three categories of crops – water intensive, drought tolerant, and cash – which are all characterised by differing levels of risk. Water intensive crops (e.g. rice) are vulnerable to rainfall shocks. Conversely drought tolerant crops (e.g. sorghum) are resistant to semi-arid conditions thereby an effective way of reducing exposure to adverse weather. Finally, cash crops (e.g.sugarcane), which cannot be directly used for household consumption as they require post-harvest processing, are generally considered to be quite profitable but also more susceptible to price fluctuations. We find that in response to a one standard deviation increase in irrigation, 13.2% of villages were more likely to report growing water intensive crops and 18.8% of villages were less likely to report cultivating crops that are drought tolerant. This provides suggestive evidence that farmers rely on predictable access to groundwater, even during the monsoon/*Kharif* season, as a form of insurance against weather shocks.

5.2 Consumption

A boost to agricultural productivity from irrigation may have important welfare implications at the village level. To capture this we conduct our analysis on a range of consumption indicators, presented in Table 5.[33] We do not detect any significant shifts in consumption per capita, the poverty rate, or night light activity (Columns 1, 2, and 4 respectively). We do however, estimate a 0.47 standard deviation increase in the household asset index for durable goods consumption (Column 3). When considering the effect independently on the main items included in this index, we find that this result is mostly driven by an increase in solid house construction (Column 1 of Table A6). The share of households that own a solid, brick and mortar, house increases by 19.8% with a one standard deviation increase in irrigation. Panel C of Figure 7 provides graphical evidence of this relationship. The share of households who own a solid house declines with groundwater depth from approximately 50% to just below 40% at the kink point, levelling off at deeper depths.

5.3 Labour

An increase in agricultural production with improved irrigation may simultaneously increase demand for labour in this sector. This effect however, may be small or even reversed if farmers switch to less labour intensive crops or replace labour activities with specialist mechanised tools such as transplanters and harvesters. Furthermore, labour supply to agriculture is likely to be influenced by market opportunities in other sectors. On-farm growth may spur production in off-farm sectors hence increasing demand for labour in those industries. Characterised by these complex interactions, the overall effect of irrigation on the sectoral allocation of labour is ambiguous.

We first consider the effect of irrigation on employment of the village population. As reported in Columns 1 to 3, Panel A of Table 6, we estimate a marginally significant decline in the total employment rate which appears to be driven by a fall in female employment. This drop may be related to the "income effect", a trend suggesting that women appear to drop out of the labour force as households become wealthier (Mehrotra et al., 2014; Mehrotra and Sinha, 2017).

Secondly, we investigate the effect of irrigation on employment in the agricultural sector presented in Columns 4 to 9 of Table 6. Agriculture is the largest employer in our sample of villages with approximately 30% of working adults reporting their primary occupation to be either cultivation (Columns 4 to 6) or manual labour (Columns 7 to 9). Irrigation however, does not appear to have any significant effects on the employment rate (Panel A) or share of the workforce (Panel B) engaged in these occupations. While we find no

[33] Beyond the direct income effect on the local population, it may be that a more productive agricultural sector spurs demand and investment in village amenities. In view of this, we investigate whether irrigation changes the probability of having access to five key services: primary school, hospital, bank, paved road, and market. However as reported in Table A5, we do not find any evidence of improvements in the availability of this infrastructure.

evidence of labour movement in or out of this sector, there may be more subtle changes occurring within the labour market. Cultivators may spend longer hours working on their farm or employ manual labour for longer periods as they cultivate more land. In order to test for this, we leverage the share of full-time workers (those that work for more than six months of the year) as the outcome of interest (Panel C). Our results suggest that there are no significant shifts on the intensive margin of work for either cultivators or labourers.

Next, we consider the effect of irrigation on labour allocation off-farm. Table 7 reports estimates on the number of persons employed across all village industries, as well as in the six largest employing industries independently. Our results suggest that irrigation does not appear to have any significant effects on employment in across these sectors.[34] Graphical evidence from Panel E of Figure 7 shows that there is no discernible change in slope in the mapping between the number of persons employed in industries and groundwater depth at the kink point. Furthermore, as demonstrated in Panel E of Figure A3 this result is tightly estimated and consistent across bandwidth down to 3 meters either side of the kink point.

Finally, we examine the possibility that irrigation may have implications on the spatial distribution of labour. Investment in tube-wells may provide villages with a comparative advantage in farming thereby pooling-in labour from neighbouring villages, especially those without access to the technology. So as to estimate this effect we consider the employment status of residents in the nearest neighbouring village from our main sample (within a maximum distance of 5km) that had no tube-wells in 2013. Using the standard regression kink specification, we report in Table 8 the impact of irrigation on agricultural sector employment for the population in these nearest non-adopting neighbours. We find evidence that in response to a one standard deviation increase in irrigation in village v, its nearest neighbour without irrigation shows a significant increase in its share of full-time agricultural labourers. This is especially so among female workers – the share of full-time female labourers increases by 24.7%.[35] In Table A8 we demonstrate that irrigation in village v has no effect on the agricultural production in its nearest neighbour without tube-wells, suggesting that the shift we estimate on full-time labour is indeed a response to higher demand for workers in relatively more agriculturally productive villages.

5.4 Demographics

Reported in Table 9, we find that irrigation leads to large changes in the village demographics. Particularly, we estimate a 38.5% increase in population density from a one

[34] Unlike the Population Census which records employment of the village population even if this takes place outside the village, the Economic Census measures the number of persons employed in village businesses even if these are not local residents. As a result, we cannot capture the employment rate or share of the workforce for workers in village industries.

[35] In Table A7 we report these results for varying distance of the nearest neighbouring village. Within 2km, presented in Columns 1 and 2, the sample is vastly reduced hence our results are not precisely estimated. However the point estimate on agricultural labourers remains very similar even at this smallest distance.

standard deviation increase in irrigation (Column 1 of Panel A). Panel F of Figure 7 provides graphical evidence of this effect. Population density falls by half with increasing groundwater depth from approximately 8 to 4 persons per hectare at the kink point and levels off at deeper depths. This considerable effect on the village population is likely due to two key pathways: (1) a more productive agricultural sector may spur in-migration, and/or (2) it provides the food supply critical in sustaining a higher fertility and/or reduced mortality.

First we consider the migration pathway by examining the share of the male population. According to the 2011 Population Census, work is the primary reason for which men migrate in India. A pooling-in of labour from outside may increase the proportion of men in the village. We do not find evidence of this shift, as presented in Column 2 of Panel B. Note however that this does not rule out in-migration of working age men, but may indicate that in the medium to long-run time frame we consider men are settling in with their families. Indeed, population density appears to increase equally across gender (Columns 2 and 3 of Panel A). Another group known to migrate for work are the Scheduled Caste. Members of these castes are among India's most economically disadvantaged groups and in 2011 represented 16% of intra-state migrants. We find evidence, shown in Columns 1 to 3 of Panel C, that irrigation causes an 8% increase in the share of this population group with similar effects for both men and women.

Second we consider the fertility and/or mortality pathway by investigating changes in the share of village population by age group. Increased fertility will lead to a higher proportion of children. Reduced mortality is likely to affect the most vulnerable, such as children and the elderly, increasing their representation in the population. Our findings indicate a significant increase of 0.89% in the share of the child (0 to 6 years) population, reported in Column 5 of Panel B. As a robustness check, we replicate this analysis using micro-data from the Socio Economic Caste Census of 2012. This dataset allows us to split the village population in 10 year age brackets. Presented in Table A9, we find evidence of a general increase across all age brackets, with the youngest population – age 1 to 10 years – showing the highest increase.

5.5 Dynamic Effects

Finally, we investigate the dynamic effects from groundwater irrigation. To do so, we leverage heterogeneity in the timing of technology adoption. As shown in Figure 1, by 2000 approximately half of our sample villages had constructed tube-wells for groundwater extraction. We therefore split our sample by villages that had built tube-wells by 2000 – for which results are presented in Table 10 – and those that invested in the technology solely after that date – with results presented in Table 11.

Panel A in both tables provides estimates on our agricultural sector indicators. The effect of irrigation in promoting agricultural production in the monsoon season is marginally higher among early adopters that have had access to the technology for at least 14 years.

In terms of changes to area cultivated, this is particularly high – an increase of 46.6% – among late adopters with access to irrigation for between 1 to 14 years. This may be due to the technology being initially adopted by wealthier villages that were already exploiting most of the available arable land. Unfortunately, we cannot test this proposition in the absence of baseline wealth measures. This would also explain our results on consumption, reported in Panels B, where we find that late adopters experience a much higher increase in their ownership of household assets.

As in the case of the overall effects from irrigation, we find no evidence of changes in aggregate employment, share of the workforce in the agricultural sector, and number of persons employed in local industries (presented in Panels C). There are however interesting differences to note on changes to the village demographics, reported in Panel D of both tables. The point estimate on the share of the child population, significant at the 5% level, is close to four times higher among early adopters compared to those who recently experienced an agricultural productivity boost. We interpret these results to suggest that increased fertility and/or reduced mortality is likely the outcome of a long to medium-run effect. On the other hand, the increased share of the Scheduled Caste population is particularly high and significant only among late adopters. Tentatively, this may indicate that in the short-run there is a large increase in demand for labour filled in by members of this group. However in the long to medium-run, an increase to the farming population may imply that they require less migrant labour.

5.6 Robustness

In this section, we examine the robustness of our results to alternative specifications and sample selection. First, in Table A10, we show that the exclusion of our geo-physical covariates do not change the results. Second, so as to account for the potential effect of temporal fluctuations in the water table we exclude villages that have experienced a large drop in their groundwater depth. Specifically, we measure the fluctuation in the maximum groundwater depth over a decade (2000-2010) and capture outliers as those in the bottom 10th percentile of the distribution (corresponding to a drop in groundwater depth by more than 4 metres). In Table A11, we demonstrate that our results are robust to excluding these villages. Finally, as explained in Section 2, the maximum theoretical threshold of a centrifugal pump is affected by altitude. As shown in Figure 4, the depth from which a perfectly efficient pump can extract water drops from 10.33 to 9.25 meters for the range of altitude covered in our main sample of villages. In order to verify that our results are not confounded with factors related to altitude, we exclude villages in the top 10th percentile of the altitude distribution (corresponding to altitudes above 600 metres). Table A12 suggests are results are robust to excluding villages at high altitudes.

6. Conclusion

A substantial literature has documented the process of economic growth across countries, overwhelmingly finding that a boost to agricultural production precedes the reallocation of labour from the agricultural sector towards the manufacturing and service industries initiating the course for industrialisation and development (Herrendorf et al., 2014). Recently, this topic has received renewed interest among micro-empirical studies to better understand the catalysts to this process, as well as how it unfolds across space.

In this paper, we analyse the effect of access to groundwater irrigation on agricultural production and the rural labour market in India. Since the 1970s, adoption of tube-wells for groundwater extraction has gradually increased making it the single largest source of irrigation. The evolution of this technology adoption over the past 50 years is well suited to empirically test the dynamic effects of productivity gains to agriculture.

To begin with, we find that irrigation significantly improves agricultural production and enables farmers to re-optimise their production strategies by cultivating more land and shifting away from drought tolerant crops. Secondly, irrigation leads to modest consumption gains, mostly with respect to durable goods. Thirdly, we find a substantial increase in population density, driven by a combination of in-migration and changes to fertility/mortality. However, there is no evidence of reallocation of labour between sectors of the rural economy. These results suggest that while villages are the root of agricultural productivity gains, industrialisation does not happen at this scale.

The strength of our empirical approach is to estimate the impact of irrigation at a high resolution across a large geographic scale. However, given that the variation in access to groundwater is identified at the village and evenly dispersed geographically, we are unable to aggregate and capture effects on the wider economy including spill-over effects to urban centres.

References

Alkire, Sabina and Suman Seth, "Identifying BPL households: A comparison of methods," *Economic and Political Weekly*, 2013, pp. 49–57.

Asher, Sam, Alison Campion, Douglas Gollin, and Paul Novosad, "The long-run development impacts of agricultural productivity gains: Evidence from irrigation canals in India," *STEG Working Paper*, 2022.

_ **and Paul Novosad**, "Rural roads and local economic development," *American Economic Review*, 2020, *110* (3), 797–823.

_ **, Tobias Lunt, Ryu Matsuura, and Paul Novosad**, "Development research at high geographic resolution: An analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 2021, *35* (4), 845–871.

Bedi, Tara, Aline Coudouel, and Kenneth Simler, *More than a pretty picture: Using poverty maps to design better policies and interventions*, World Bank Publications, 2007.

Blakeslee, David, Aaditya Dar, Ram Fishman, Samreen Malik, Heitor S Pellegrina, and Karan Singh Bagavathinathan, "Irrigation and the spatial pattern of local economic development in India," *Journal of Development Economics*, 2023, *161*, 102997.

_ **, Ram Fishman, and Veena Srinivasan**, "Way down in the hole: Adaptation to long-term water loss in rural India," *American Economic Review*, 2020, *110* (1), 200–224.

Bleakley, Hoyt and Jeffrey Lin, "Portage and path dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural productivity and structural transformation: Evidence from Brazil," *American Economic Review*, 2016, *106* (6), 1320–65.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, "Robust data-driven inference in the regression-discontinuity design," *The Stata Journal*, 2014, *14* (4), 909–946.

Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei, "The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in Missouri, 2003-2013," *American Economic Review*, 2015, *105* (5), 126–30.

__ , **David S Lee, Zhuan Pei, and Andrea Weber**, "Inference on causal effects in a generalized regression kink design," *Econometrica*, 2015, *83* (6), 2453–2483.

Dubash, Navroz K, "The electricity-groundwater conundrum: Case for a political solution to a political problem," *Economic and Political Weekly*, 2007, pp. 45–55.

Duflo, Esther and Rohini Pande, "Dams," *The Quarterly Journal of Economics*, 2007, *122* (2), 601–646.

Elbers, Chris, Jean O Lanjouw, and Peter Lanjouw, "Micro-level estimation of poverty and inequality," *Econometrica*, 2003, *71* (1), 355–364.

Elsey, Jim, "How to define and measure centrifugal pump efficiency," *Pumps and Systems*, 2020.

Elvidge, Christopher, Kimberly Baugh, Vinita Hobson, Eric Kihn, Herbert Kroehl, Ethan Davis, and David Cocero, "Satellite inventory of human settlements using nocturnal radiation emissions: A contribution for the global toolchest," *Global Change Biology*, 1997, *3* (5), 387–395.

Emerick, Kyle, "Agricultural productivity and the sectoral reallocation of labor in rural India," *Journal of Development Economics*, 2018, *135*, 488–503.

Faber, Tom E, *Fluid dynamics for physicists*, Cambridge university press, 1995.

Famiglietti, James S, "The global groundwater crisis," *Nature Climate Change*, 2014, *4* (11), 945–948.

Faurès, Jean-Marc, Jippe Hoogeveen, and Jelle Bruinsma, "The FAO irrigated area forecast for 2030," *FAO, Rome, Italy*, 2002, pp. 1–14.

Foster, Andrew D and Mark R Rosenzweig, "Technical change and human-capital returns and investments: evidence from the green revolution," *The American economic review*, 1996, pp. 931–953.

__ **and** __ , "Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000," *Economic Development and Cultural Change*, 2004, *52* (3), 509–542.

__ **and** __ , "Technological change and the distribution of schooling: evidence from green-revolution India," *Journal of Development Economics*, 2004, *74* (1), 87–111.

Funk, Chris C, Pete J Peterson, Martin F Landsfeld, Diego H Pedreros, James P Verdin, James D Rowland, Bo E Romero, Gregory J Husak, Joel C Michaelsen, Andrew P Verdin et al., "A quasi-global precipitation time series for drought monitoring," *US Geological Survey data series*, 2014, *832* (4), 1–12.

Funk, Chris, Pete Peterson, Seth Peterson, Shraddhanand Shukla, Frank Davenport, Joel Michaelsen, Kenneth R Knapp, Martin Landsfeld, Gregory Husak, Laura Harrison et al., "A high-resolution 1983–2016 T max climate data record based on infrared temperatures and stations by the Climate Hazard Center," *Journal of Climate*, 2019, *32* (17), 5639–5658.

Gao, Xiang, Alfredo R Huete, Wenge Ni, and Tomoaki Miura, "Optical–biophysical relationships of vegetation spectra without background contamination," *Remote Sensing of Environment*, 2000, *74* (3), 609–620.

Garmann, Sebastian, "The causal effect of coalition governments on fiscal policies: Evidence from a regression kink design," *Applied Economics*, 2014, *46* (36), 4490–4507.

Gollin, Douglas, Casper Worm Hansen, and Asger Mose Wingender, "Two blades of grass: The impact of the green revolution," *Journal of Political Economy*, 2021, *129* (8), 2344–2384.

_ , Stephen Parente, and Richard Rogerson, "The role of agriculture in development," *American Economic Review*, 2002, *92* (2), 160–164.

Harari, Mariaflavia, "Cities in bad shape: Urban geometry in India," *American Economic Review*, 2020, *110* (8), 2377–2421.

Henderson, Vernon, Adam Storeygard, and David N Weil, "A bright idea for measuring economic growth," *American Economic Review*, 2011, *101* (3), 194–99.

Hentschel, Jesko, Jean Olson Lanjouw, Peter Lanjouw, and Javier Poggi, "Combining census and survey data to trace the spatial dimensions of poverty: A case study of Ecuador," *The World Bank Economic Review*, 2000, *14* (1), 147–165.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi, "Growth and structural transformation," in "Handbook of economic growth," Vol. 2, Elsevier, 2014, pp. 855–941.

Hodler, Roland and Paul A Raschky, "Regional favoritism," *The Quarterly Journal of Economics*, 2014, *129* (2), 995–1033.

Hornbeck, Richard and Pinar Keskin, "The historically evolving impact of the ogallala aquifer: Agricultural adaptation to groundwater and drought," *American Economic Journal: Applied Economics*, 2014, *6* (1), 190–219.

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

Imbens, Guido W and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of econometrics*, 2008, *142* (2), 615–635.

Jain, Rajni, Prabhat Kishore, and Dhirendra Kumar Singh, "Irrigation in India: Status, challenges and options," *Journal of Soil and Water Conservation*, 2019, *18* (4), 354–363.

Kouadio, Louis, Nathaniel K Newlands, Andrew Davidson, Yinsuo Zhang, and Aston Chipanshi, "Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale," *Remote Sensing*, 2014, *6* (10), 10193–10214.

Labus, MP, GA Nielsen, RL Lawrence, R Engel, and DS Long, "Wheat yield estimates using multi-temporal NDVI satellite imagery," *International Journal of Remote Sensing*, 2002, *23* (20), 4169–4180.

Landais, Camille, "Assessing the welfare effects of unemployment benefits using the regression kink design," *American Economic Journal: Economic Policy*, 2015, *7* (4), 243–78.

Manring, Noah, *Fluid power pumps and motors: Analysis, design and control*, McGraw Hill Professional, 2013.

Matsuyama, Kiminori, "Agricultural productivity, comparative advantage, and economic growth," *Journal of Economic Theory*, 1992, *58* (2), 317–334.

McArthur, John W and Gordon C McCord, "Fertilizing growth: Agricultural inputs and their effects in economic development," *Journal of Development Economics*, 2017, *127*, 133–152.

McKenzie, David J, "Measuring inequality with asset indicators," *Journal of Population Economics*, 2005, *18* (2), 229–260.

Mearns, Robin, *Access to land in rural India: policy issues and options*, Vol. 2123, World Bank Publications, 1999.

Mehrotra, Santosh and Sharmistha Sinha, "Explaining falling female employment during a high growth period," *Economic and Political Weekly*, 2017, *52* (39), 54–62.

_ , Jajati Parida, Sharmistha Sinha, and Ankita Gandhi, "Explaining employment trends in the Indian economy: 1993-94 to 2011-12," *Economic and Political Weekly*, 2014, pp. 49–57.

Mellander, Charlotta, José Lobo, Kevin Stolarick, and Zara Matheson, "Night-time light data: A good proxy measure for economic activity?," *PloS one*, 2015, *10* (10), e0139779.

Minale, Luigi, "Agricultural productivity shocks, labour reallocation and rural–urban migration in China," *Journal of Economic Geography*, 2018, *18* (4), 795–821.

Mukherji, Aditi, "Evolution of irrigation sector," *Economic and Political Weekly*, 2016, pp. 44–47.

Ngai, L Rachel and Christopher A Pissarides, "Structural change in a multisector model of growth," *American Economic Review*, 2007, *97* (1), 429–443.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher Taber, "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2010, *2* (2), 185–215.

Nurkse, Ragnar, "Problems of capital formation in underdeveloped countries," 1953.

Rajan, Abhishek and Shilp Verma, "Evolving nature of India's Irrigation Economy," *Water Policy Research Highlight*, 2017.

Rasmussen, Michael S, "Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability," *International Journal of Remote Sensing*, 1997, *18* (5), 1059–1077.

Robinson, Austin, "The changing structure of the British economy," *The Economic Journal*, 1954, *64* (255), 443–461.

Ryan, Nicholas and Anant Sudarshan, "Rationing the commons," *Journal of Political Economy*, 2022, *130* (1), 210–257.

Sarsons, Heather, "Rainfall and conflict: A cautionary tale," *Journal of Development Economics*, 2015, *115*, 62–72.

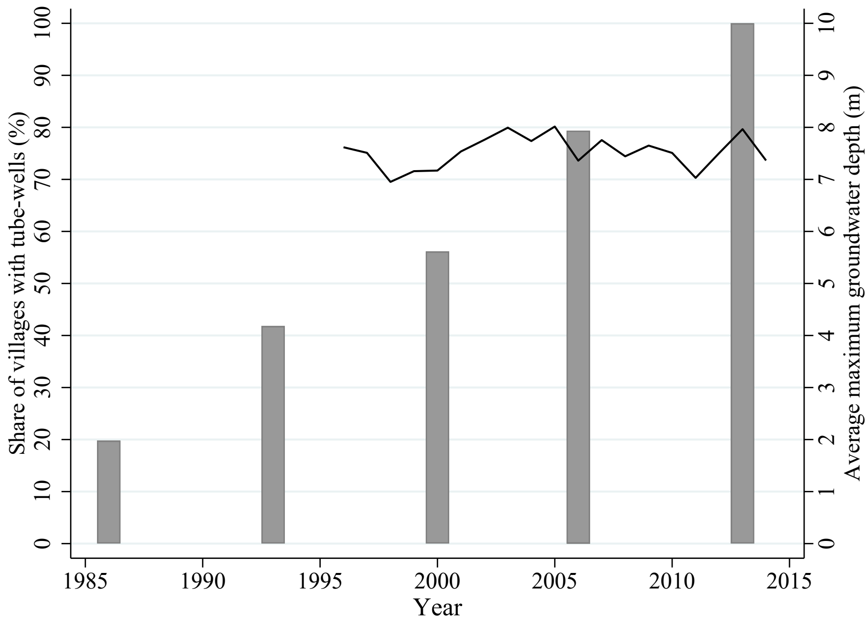
Sekhri, Sheetal, "Wells, water, and welfare: The impact of access to groundwater on rural poverty and conflict," *American Economic Journal: Applied Economics*, 2014, *6* (3), 76–102.

Shah, Tushaar, Mark Giordano, and Aditi Mukherji, "Political economy of the energy-groundwater nexus in India: exploring issues and assessing policy options," *Hydrogeology Journal*, 2012, *20* (5), 995.

Simonsen, Marianne, Lars Skipper, and Niels Skipper, "Price sensitivity of demand for prescription drugs: Exploiting a regression kink design," *Journal of Applied Econometrics*, 2016, *31* (2), 320–337.

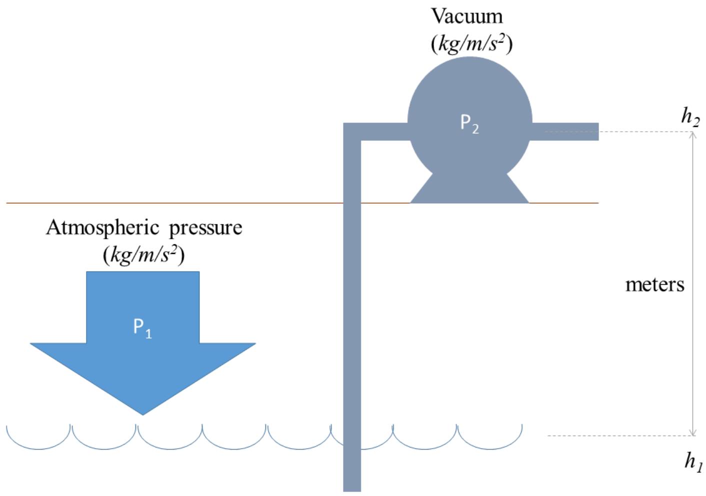
Wardlow, Brian D and Stephen L Egbert, "A comparison of MODIS 250-m EVI and NDVI data for crop mapping: A case study for Southwest Kansas," *International Journal of Remote Sensing*, 2010, *31* (3), 805–830.

Figure 1: Tube-well construction and groundwater depth over time



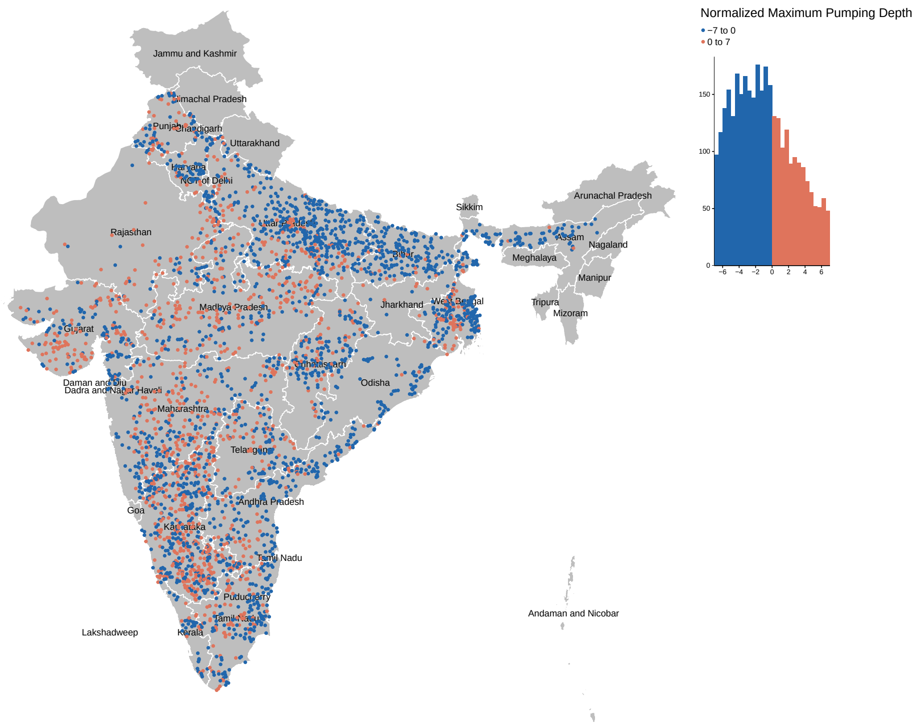
Notes: The percentage share of villages with tube-wells is represented by the bar graph with its axis on the left. Data on tube-well construction is obtained from the Minor Irrigation (MI) Censuses conducted every seven years since 1986. Annual maximum groundwater depth is represented by the line graph with its axis on the right. Data on groundwater depth comes from the Central Ground Water Board (CGWB) which has been monitoring wells across the country since 1996. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample).

Figure 2: Pumping mechanism of a centrifugal pump



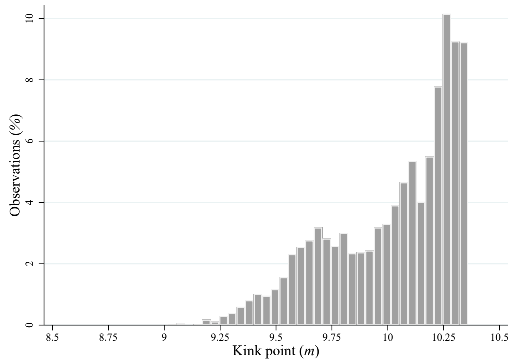
Notes: Extraction of water from a tube-well using a centrifugal pump can be described by Bernoulli's principle of fluid dynamics (Equation 1). Assuming constant flow velocity Equation 1 can be re-written in the form of Equation 2: $h_2 - h_1 = \frac{P_1 - P_2}{\rho g}$, where P_1 and P_2 refer to pressure from the water table and the pump mechanism respectively, ρ is the density of water (997 kg/m^3), and g the gravitational force (9.81 m/s^2). The maximum possible pressure differential is achieved through a perfect vacuum (P_2=0 $kg/m/s^2$) in the pumping mechanism. Under this ideal condition and atmospheric pressure at sea-level (P_1=101,325 $kg/m/s^2$) the maximum depth from which water can be extracted, calculated as the difference between h_1 and h_2, is 10.33 meters. This represents the maximum theoretical threshold which a centrifugal pump can achieve.

Figure 3: Sample village location and groundwater depth

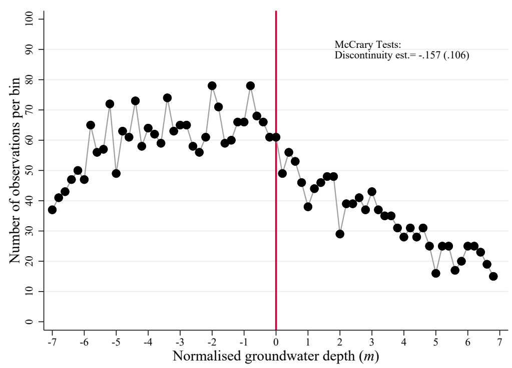


Notes: The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Each circle on the map represents a village in this sample. Red circles correspond to villages with a groundwater depth deeper than the kink point. Blue points are villages with a groundwater depth shallower than the kink point. The kink point of a village is calculated using Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude.

Figure 4: Distribution of the assignment variable



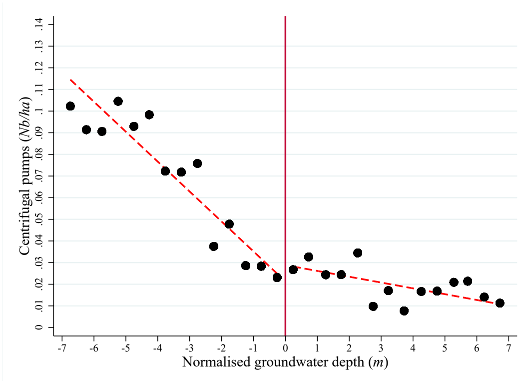
Panel A



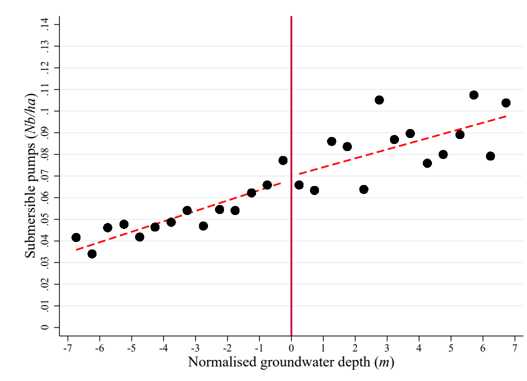
Panel B

Notes: The kink point of a village is calculated using Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude. Panel A shows the distribution of the kink point for villages in our sample. Panel B plots the number of observations in each bin for groundwater depth normalised at the threshold. A *fuzzy* RK design requires for the conditional density of the assignment variable, given the unobserved error in the outcome, to be continuously differentiable at the kink point. The McCrary test, reported in Panel B, provides an additional validation by estimating the log change in height between bins at that point. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample).

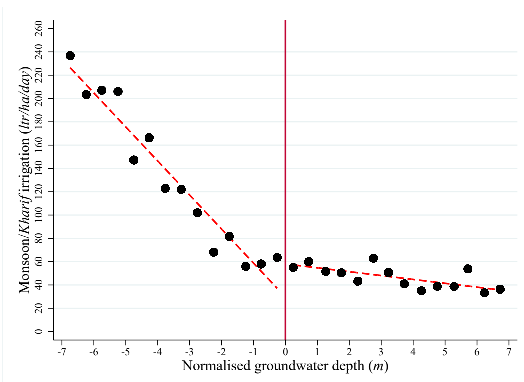
Figure 5: Deterministic relation between groundwater depth, pump adoption, and irrigation



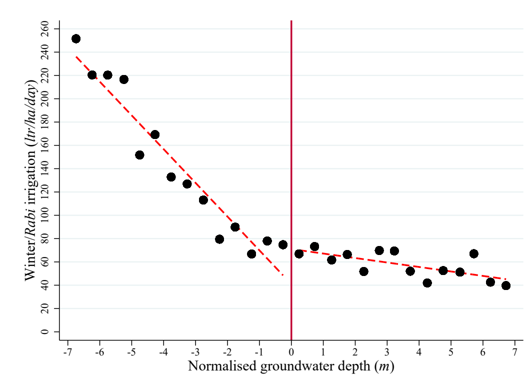
Panel A



Panel B



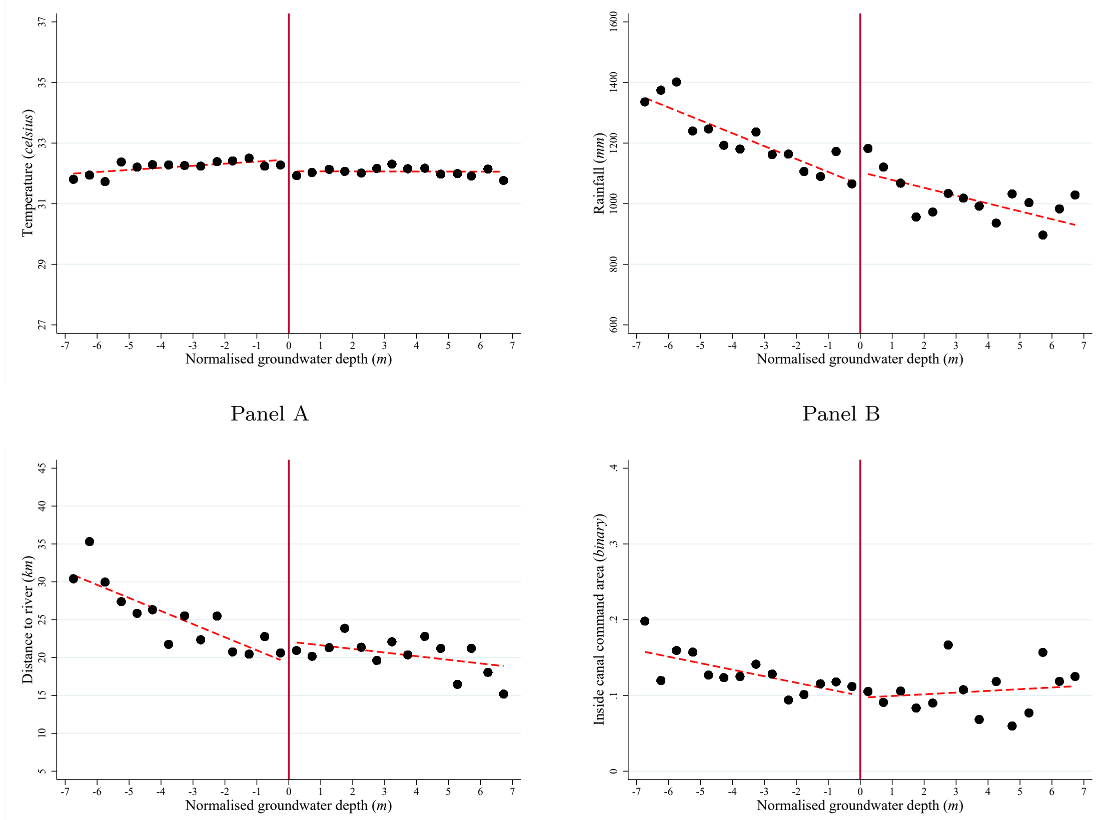
Panel C



Panel D

Notes: The x-axis in each panel represents our assignment variable, groundwater depth. This variable is normalised around the kink point of the village. The kink point is calculated using Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude. Points to the right of zero correspond to depths deeper than the kink point, while those left of zero are shallower. Each panel reports results on the deterministic relation between our assignment variable and measures of pump adoption and irrigation. Each panel shows the mean values of the variable of interest in each bin of the assignment variable. The bin size is 0.5. The red dashed lines display predicted values of the regressions in the linear case allowing for a discontinuous shift at the kink point. Formal estimates of the kink using the *fuzzy* RK specification are reported in Table 2. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample).

Figure 6: Deterministic relation between groundwater depth and geo-physical covariates

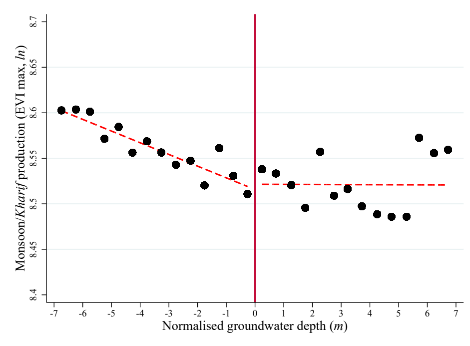


Panel A Panel B

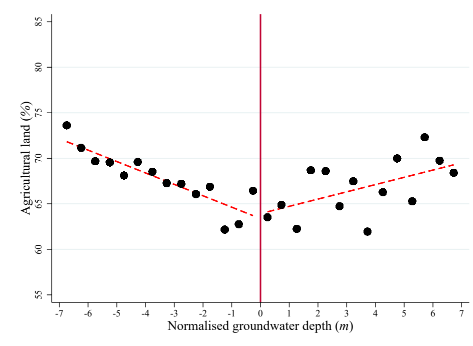
Panel C Panel D

Notes: The x-axis in each panel represents our assignment variable, groundwater depth. This variable is normalised around the kink point of the village. The kink point is calculated using Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude. Points to the right of zero correspond to depths deeper than the kink point, while those left of zero are shallower. Each panel reports results on the deterministic relation between our assignment variable and covariates. Each panel shows the mean values of the covariate in each bin of the assignment variable. The bin size is 0.5. The red dashed lines display predicted values of the regressions in the linear case allowing for a discontinuous shift at the kink point. Formal estimates of the kink using the *fuzzy* RK specification are reported in Table 2. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample).

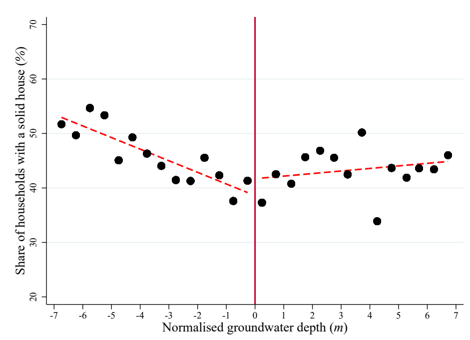
Figure 7: Deterministic relation between groundwater depth and outcomes



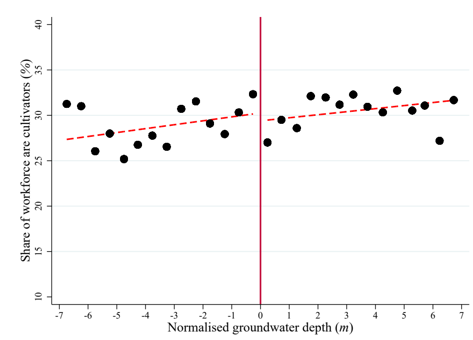
Panel A



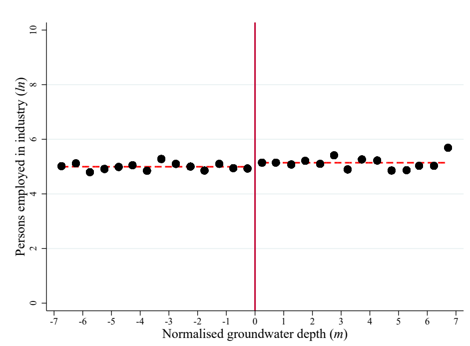
Panel B



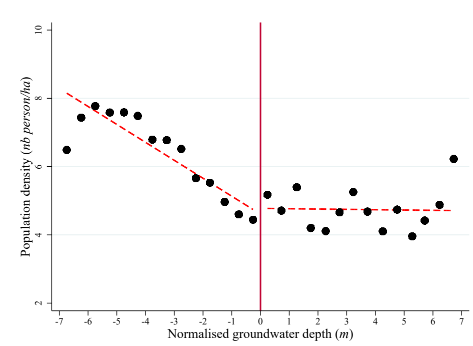
Panel C



Panel D



Panel E



Panel F

Notes: The x-axis in each panel represents our assignment variable, groundwater depth. This variable is normalised around the kink point of the village. The kink point is calculated using Bernoulli's principle of fluid dynamics described in Equation 2 assuming 100% pump efficiency and atmospheric pressure adjusted for village altitude. Points to the right of zero correspond to depths deeper than the kink point, while those left of zero are shallower. Each panel reports results on the deterministic relation between our assignment variable and a selection of our outcome variables. Each panel shows the mean values of the outcome of interest in each bin of the assignment variable. The bin size is 0.5. The red dashed lines display predicted values of the regressions in the linear case allowing for a discontinuous shift at the kink point. Formal estimates of the kink using the *fuzzy* RK specification are reported in Tables 4, A6, 6, 7, and 9 for Panels A to F respectively. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample).

Table 1: Descriptive statistics of sample

	Mean (1)	SD (2)	N (3)	Source (Year) (4)
Panel A: Irrigation				
Agricultural area irrigated by tube-wells (%)	76.425	42.347	3327	PC (2011)
Monsoon/*Kharif* irrigation (*ltr/ha/day*)	97.677	164.205	3327	MIC (2013)
Winter/*Rabi* irrigation (*ltr/ha/day*)	108.561	170.330	3327	MIC (2013)
Tube-wells (*nb/ha*)	0.123	0.165	3327	MIC (2013)
Centrifugal pumps (*nb/ha*)	0.050	0.109	3327	MIC (2013)
Maximum groundwater depth (*m*)	8.964	3.468	3327	CGWB (2010-13)[a]
Panel B: Agriculture				
Landholding size (*ha*)	3.541	5.967	2349	SECC (2012)
Share of HHs with mechanised equipment (%)	5.007	8.702	2349	SECC (2012)
Panel C: Consumption				
Per capita consumption ('000 *Rs./annum*)	18.439	4.547	3327	SECC (2012)
Share of HHs that are BPL[b] (%)	28.611	17.304	3327	SECC (2012)
Share of HHs who own a solid house (%)	44.444	29.047	2349	SECC (2012)
Panel D: Labour				
Share of workforce are cultivators (%)	13.027	9.669	3327	PC (2011)
Share of workforce are agricultural labourers (%)	18.198	11.399	3327	PC (2011)
Persons employed in village businesses (*nb*)	435.149	697.281	3327	EC (2013)
Panel E: Demographics				
Population (*nb*)	4029.857	4030.772	3327	PC (2011)
Share of population from scheduled castes (%)	20.004	16.366	3327	PC (2011)
Panel F: Covariates				
Temperature (*celsius*)	32.171	1.735	3327	CHIRTS (2010-13)[a]
Rainfall (*mm*)	1139.756	512.546	3327	CHIRPS (2010-13)[a]
Distance to nearest river (*km*)	23.400	25.486	3327	PC (2011)

Notes: This table presents summary statistics of the sample captured between 2011 to 2013 depending on the source of data. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample).
[a]Calculated as a three year average between 2010 to 2013.
[b]Poverty line is set at Rs.31/day.

Table 2: Estimated kink in the deterministic relation of groundwater depth with pump adoption, irrigation, and covariates

	Pump adoption		Irrigation			Covariates			
	Centrifugal	Submersible	Average	Monsoon/ *Kharif*	Winter/ *Rabi*	Temperature	Rainfall	Distance to river	Inside canal command area
	(nb/ha)	*(nb/ha)*	*(standardised)*	*(standardised)*	*(standardised)*	*(celsius)*	*(mm)*	*(km)*	*(binary)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
δ_2	0.003**	0.001	0.107***	0.108***	0.102***	-0.025	7.156	0.539	0.009
	(0.001)	(0.002)	(0.014)	(0.014)	(0.014)	(0.025)	(7.217)	(0.412)	(0.006)
Mean	0.050	0.063	0.000	0.000	0.000	32.171	1139.756	23.400	0.119
SD	0.109	0.112	1.000	1.000	1.000	1.735	512.546	25.486	0.324
N	3327	3327	3327	3327	3327	3327	3327	3327	3327

Notes: This table presents estimates on the effect of groundwater depth on pump adoption, irrigation, and covariates. δ_2 is the estimated change in slope of the assignment function at the kink point (Equation 3). Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal (Column 1) and submersible (Column 2) pumps. Irrigation is measured as water input in litres and standardised. We report irrigation as an average over the year (Column 3), as well as independently for the Monsoon/*Kharif* (Column 4) and the dry Winter/*Rabi* season (Column 5). We consider four geo-physical covariates reported in Columns 6 to 9 respectively: temperature (measured as a three year average, 2010-13, of the maximum monthly temperature recorded in degrees Celsius), rainfall (measured as a three year average, 2010-13, of the total annual rainfall recorded in millimetres), distance to the nearest river (in kilometres), and a binary indicator for whether the village has tube-wells inside the command area of a canal network. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation is reported for the full sample. Each regression includes state dummies and covariates; with the covariate of interest omitted from the vector of controls (see Section 4 for specification details). Heteroskedasticity robust standard errors are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3: Balance of outcome variables pre-treatment for villages with tube-wells built post 2000

	Full sample (1)	Deep ($w > k$) (2)	Shallow ($w \leq k$) (3)	Difference in means (4)	p-value on difference (5)	RK estimate (6)	p-value on RK estimate (7)
Panel A: Agriculture							
Area irrigated by tube-wells (%)	3.424	2.921	3.713	0.792	0.35	2.195	0.67
Monsoon/*Kharif* production (EVI max, *ln*)	8.944	8.929	8.953	0.024	0.00	0.026	0.66
Winter/*Rabi* production (EVI max, *ln*)	8.457	8.425	8.474	0.049	0.00	-0.055	0.48
Land (*ln*)	5.937	6.051	5.868	-0.183	0.03	0.710	0.28
Panel B: Consumption							
HHs with income above Rs.250/month (%)	80.253	81.659	79.315	-2.344	0.19	-9.476	0.54
HHs who own land (%)	57.048	60.940	54.626	-6.314	0.00	-20.506	0.15
Night light (*ln*)	1.736	1.800	1.700	-0.100	0.00	0.285	0.24
Panel C: Demographics							
Population (*ln*)	7.612	7.600	7.620	0.021	0.76	0.830	0.13
Population from scheduled castes (%)	17.521	16.728	18.000	1.272	0.21	14.248	0.12
N	1459	532	927				

Notes: The table presents summary statistics and balance tests pre-treatment for villages with tube-wells built after 2000. Data on demographics are obtained from the 2001 Population Census of India, consumption indicators are from the 2002 Below Poverty Line Census, irrigation variables come from the Third Minor Irrigation Census of 2000, agricultural production (EVI max) and night light (mean) are taken from satellite imagery captured for the year 2000. Columns 1 to 3 show the unconditional mean for all villages, villages with groundwater depths deeper than the kink point, and villages with groundwater depths shallower than the kink point respectively. Column 4 presents the difference in means between Columns 2 and 3. Column 5 shows the p-value for the difference in means. Column 6 reports the regression kink estimates capturing the effect of groundwater irrigation on each variable. The specification includes state dummies (see Section 4 for details). Finally Column 7 presents the p-value for the regression kink estimates. The sample consists of villages with tube-wells built after 2000 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample).

Table 4: Impact of irrigation on agriculture

	Production		Inputs			Crop choice		
	Monsoon/ *Kharif* (EVI max, *ln*)	Winter/ *Rabi* (EVI max, *ln*)	Agricultural land (%)	Water-saving technology (%)	Mechanised equipment (%)	Water intensive (*binary*)	Drought tolerant (*binary*)	Cash (*binary*)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Irrigation	0.077***	-0.005	18.711***	-6.051*	-1.740	0.132*	-0.188**	0.012
(*standardised*)	(0.030)	(0.026)	(4.579)	(3.311)	(1.869)	(0.079)	(0.083)	(0.077)
Mean	4609.952	4865.387	67.034	4.753	5.007	0.689	0.343	0.227
SD	949.797	1041.578	24.487	18.671	8.702	0.463	0.475	0.419
N	3327	3327	3327	3327	2349	2700	2700	2700

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on agricultural output and production choices. Irrigation is measured as *litres/ha/day* and standardised. Using EVI, an index of vegetation cover from satellite imagery, we proxy for agricultural production by taking the maximum value of the index (log transformed) in both the Monsoon/*Kharif* (Column 1) and the dry Winter/*Rabi* season (Column 2) of 2013. In Columns 3 to 5 we consider adoption of three inputs respectively: agricultural land (percentage share of village area used for agricultural purposes), water-saving technology (percentage share of tube-wells which are adapted to water-saving mechanisms such as drips and sprinklers), and mechanisation (percentage share of households who own mechanised farm equipment such as tractors, harvesters etc.). Additionally, we also report results on three binary measures of crop choice reported in Columns 6 to 8 respectively: does a village grow water intensive crops (sugarcane, cotton, and rice), drought tolerant crops (millet, sorghum, maize, pigeon pea, and groundnut), and cash crops (sugarcane, oilseed, cotton, and tobacco). The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample, and in the case of production on the level form of the variables. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 5: Impact of irrigation on consumption

	Consumption per capita	Poverty rate	Household assets	Night light
	(*ln*)	(*share*)	(*index*)	(*ln*)
	(1)	(2)	(3)	(4)
Irrigation	0.027	-0.037	0.470**	0.100
(*standardised*)	(0.034)	(0.026)	(0.221)	(0.084)
Mean	18.662	0.290	0.409	7.307
SD	4.762	0.178	1.000	5.200
N	3327	3327	2349	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on consumption. Irrigation is measured as *litres/ha/day* and standardised. Column 1 reports results on the imputed consumption per capita (log transformed). Column 2 shows estimates on the imputed share of the population living below the poverty line (poverty line is set at Rs.31/day). Column 3 shows the household asset ownership index calculated as the village level average of the primary component of indicator variables for all household assets captured in the Socio Economic Caste Census of 2012. Finally, using satellite imagery, Column 4 captures the average night light in 2013. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample, and in the case of night light and consumption on the level form of the variables. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis, except for consumption and poverty which report bootstrapped standard errors. * significant at 10% ** significant at 5% *** significant at 1%.

Table 6: Impact of irrigation on aggregate and agricultural sector employment

	Total			Cultivators			Labourers		
	Person (1)	Male (2)	Female (3)	Person (4)	Male (5)	Female (6)	Person (7)	Male (8)	Female (9)
Panel A: Share of population employed (%)									
Irrigation	-2.576*	-0.859	-4.285*	-1.145	-0.358	-2.022	0.426	0.640	0.251
(*standardised*)	(1.445)	(0.872)	(2.495)	(1.475)	(1.772)	(1.517)	(1.741)	(1.756)	(2.091)
Mean	43.784	55.232	31.706	13.027	18.303	7.480	18.198	19.188	17.093
SD	10.526	7.016	17.376	9.669	11.345	9.964	11.399	11.172	14.057
Panel B: Share of workforce (%)									
Irrigation	-	-	-	-0.312	0.365	-2.064	2.780	1.628	5.597
(*standardised*)	(-)	(-)	(-)	(2.939)	(3.088)	(3.212)	(3.215)	(3.013)	(4.227)
Mean	-	-	-	29.393	33.041	21.238	40.293	34.663	49.859
SD	-	-	-	18.856	19.589	20.592	20.671	19.358	26.284
Panel C: Share of full-time workers (%)									
Irrigation	4.839	3.734	4.261	4.457	2.069	4.328	3.501	2.842	6.568
(*standardised*)	(3.298)	(2.878)	(4.480)	(2.843)	(2.468)	(5.046)	(4.723)	(4.672)	(5.339)
Mean	75.369	82.137	61.180	86.884	91.027	69.532	64.495	70.671	54.374
SD	20.638	18.258	28.515	17.917	15.904	31.141	29.690	29.227	33.989
N	3327	3327	3327	3327	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on aggregate employment as well as within the agricultural sector. Irrigation is measured as *litres/ha/day* and standardised. Panel A reports results on the percentage share of the population employed, calculated as the ratio of those employed to the total working age population. Panel B reports results on the percentage share of the workforce, calculated as the ratio of those employed to the total workforce. Panel C reports results on the percentage share of full-time workers (those that work for more than 6 months of the year), calculated as the ratio of full-time workers to the total workforce. Alongside total employment we consider two specific occupational categories in agriculture: cultivators are those who cultivate their own land, and labourers are those who work for a daily wage. Furthermore, we disaggregate each of our categories by gender. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 7: Impact of irrigation on industry sector employment

	All	Agro-processing	Livestock	Education	Manufacturing	Services	Forestry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(ln)	(ln)	(ln)	(ln)	(ln)	(ln)	(ln)
Irrigation	0.132	-0.079	-0.166	-0.066	0.472	0.387	-0.093
($standardised$)	(0.248)	(0.178)	(0.316)	(0.238)	(0.299)	(0.276)	(0.110)
Mean	435.149	4.357	100.878	36.186	81.043	183.974	1.228
SD	697.281	16.939	235.447	55.302	176.170	316.905	7.174
N	3327	3327	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on employment within village industries. Irrigation is measured as $litres/ha/day$ and standardised. Results are reported on the number of persons employed (log transformed). Column 1 calculates employment on aggregate across all village industries. Columns 2 to 7 refer to each of the following largest employing sectors respectively: agro-processing, livestock, education, manufacturing, services, and forestry. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation are reported on the level form of the variables for the full sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 8: Impact of irrigation on the spatial distribution of agricultural sector employment

	Cultivators			Labourers		
	Person (1)	Male (2)	Female (3)	Person (4)	Male (5)	Female (6)
Panel A: Share of population employed (%)						
Irrigation	-1.605	-1.124	-1.170	1.897	3.042	0.751
(*standardised*)	(2.481)	(2.892)	(2.498)	(2.830)	(2.988)	(3.134)
Mean	13.573	18.542	8.360	17.267	18.272	16.179
SD	13.027	15.305	13.342	15.007	15.396	17.227
Panel B: Share of workforce (%)						
Irrigation	-2.928	-2.139	-0.412	4.871	4.309	6.981
(*standardised*)	(4.767)	(4.919)	(4.858)	(5.097)	(4.853)	(6.203)
Mean	29.893	33.263	21.601	37.065	32.721	43.861
SD	25.020	26.167	25.851	27.328	25.968	32.995
Panel C: Share of full-time workers (%)						
Irrigation	4.263	2.419	10.443	17.910**	16.098**	24.697***
(*standardised*)	(6.550)	(6.570)	(7.973)	(7.553)	(7.622)	(8.116)
Mean	73.839	77.374	55.057	53.745	59.031	44.110
SD	35.235	35.501	41.459	38.062	38.998	40.009
N	2287	2287	2287	2287	2287	2287

Notes: This table presents *fuzzy* RK estimates on the spatial distribution effect of irrigation on agricultural sector employment. These effects are measured for the nearest neighbouring village without access to groundwater irrigation. Irrigation is measured as $litres/ha/day$ and standardised. Panel A reports results on the percentage share of the population employed, calculated as the ratio of those employed to the total working age population. Panel B reports results on the percentage share of the workforce, calculated as the ratio of those employed to the total workforce. Panel C reports results on the percentage share of full-time workers (those that work for more than 6 months of the year), calculated as the ratio of full-time workers to the total workforce. We consider two specific occupational categories in agriculture: cultivators are those who cultivate their own land, and labourers are those who work for a daily wage. Furthermore, we disaggregate each of our categories by gender. The sample consists of villages without tube-wells in 2013 that are the nearest neighbour within a 5km distance from our main sample of villages (villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point, see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the nearest neighbour sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 9: Impact of irrigation on village demographics

	Persons (1)	Male (2)	Female (3)	Adult (4)	Child (5)
Panel A: Population density (*ln*)					
Irrigation	0.385***	0.382***	0.390***	0.344***	0.466***
(*standardised*)	(0.128)	(0.128)	(0.129)	(0.130)	(0.133)
Mean	5.761	2.956	2.814	3.757	0.784
SD	5.556	2.879	2.723	3.635	0.836
Panel B: Share of the total population (%)					
Irrigation	-	-0.206	-	-	0.885**
(*standardised*)	(-)	(0.289)	(-)	(-)	(0.446)
Mean	-	51.162	-	-	13.165
SD	-	2.066	-	-	3.294
Panel C: Share of scheduled caste population (%)					
Irrigation	8.249***	8.400***	8.079***	-	-
(*standardised*)	(2.886)	(2.899)	(2.881)	(-)	(-)
Mean	20.004	19.942	20.068	-	-
SD	16.366	16.363	16.441	-	-
N	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on village demographics. Irrigation is measured as *litres/ha/day* and standardised. Panel A presents results on population density, calculated as the ratio of the population to village area (log transformed). We present estimates for the total population in Column 1, as well as disaggregated by gender (Columns 2 and 3 for male and female respectively) and age (Columns 4 and 5 for adults, 18+ years, and child, 0-6 years, respectively). Panel B reports results on the percentage share of the male and child population, calculated as the ratio of that population to the total population. Panel C presents results on the percentage share of the Scheduled Caste population, calculated as the ratio of that population to the total population. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample, and in the case of population density on the level form of the variables. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 10: Impact of irrigation for villages with tube-wells built post 2000

Panel A: Agriculture	Monsoon/*Kharif* production (EVI max, *ln*)	Winter/*Rabi* production (EVI max, *ln*)	Agricultural land (%)	Water intensive (*binary*)	Drought tolerant (*binary*)	
Irrigation	0.064	-0.052	46.634***	0.153	-0.320	
(*standardised*)	(0.073)	(0.067)	(15.623)	(0.174)	(0.206)	
Panel B: Consumption	Household assets (*index*)	Night light (*ln*)				
Irrigation	1.634***	0.069				
(*standardised*)	(0.618)	(0.208)				
Panel C: Labour	Population employed (%)	Cultivators (%)	Manual labourers (%)	Education (*ln*)	Manufacturing (*ln*)	Services (*ln*)
Irrigation	-4.753	1.522	0.163	-0.110	1.549*	1.175
(*standardised*)	(3.672)	(7.823)	(8.135)	(0.557)	(0.812)	(0.718)
Panel D: Demographics	Population density (*ln*)	Male share (%)	Child share (%)	SC share (%)		
Irrigation	0.572*	0.179	0.380	21.464**		
(*standardised*)	(0.328)	(0.711)	(1.147)	(8.532)		
N	1459	1459	1459	1459	1459	1459

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on our key outcomes for villages with tube-wells built after 2000. Irrigation is measured as *litres/ha/day* and standardised. Panel A reports results on the agricultural sector including: production in the Monsoon/*Kharif* and Winter/*Rabi* season (EVI max log transformed), share of village area used for agriculture, whether the village grows water intensive and drought tolerant crops. Panel B reports results on consumption including: household asset ownership index and log of average night light luminosity. Panel C reports results on labour allocation, including: aggregate employment rate, share of the workforce in agriculture, and number of persons employed in village industries (log transformed). Panel D reports results on village demographics including: log of population density, population share of adults, children, and Scheduled Caste. The sample consists of villages with tube-wells built between 2000 to 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 11: Impact of irrigation for villages with tube-wells built before 2001

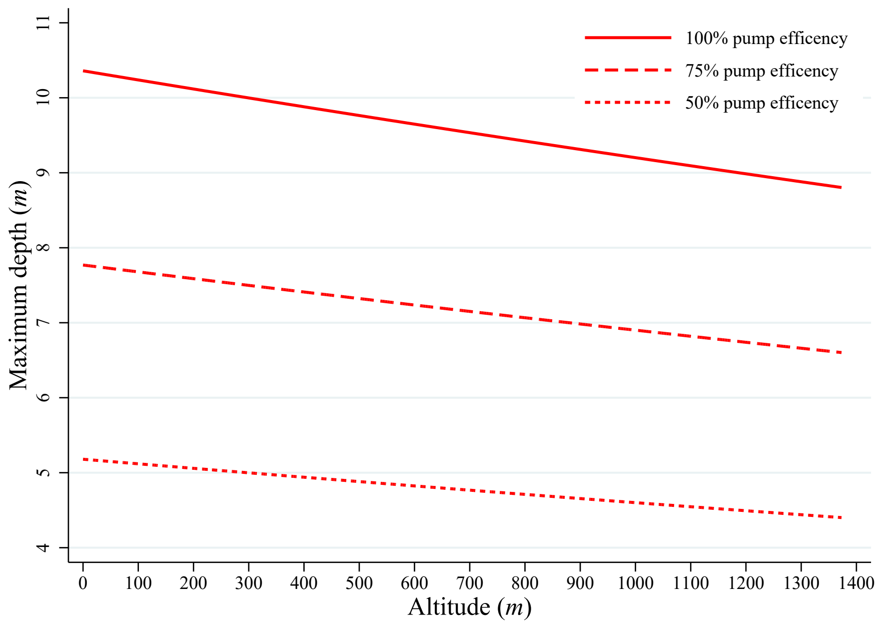
Panel A: Agriculture						
	Monsoon/*Kharif* production (EVI max, *ln*)	Winter/*Rabi* production (EVI max, *ln*)	Agricultural land (%)	Water intensive (*binary*)	Drought tolerant (*binary*)	
Irrigation	0.087***	0.015	7.552*	0.127	-0.116	
(*standardised*)	(0.032)	(0.027)	(3.873)	(0.083)	(0.079)	
Panel B: Consumption						
	Household assets (*index*)	Night light (*ln*)				
Irrigation	0.034	0.089				
(*standardised*)	(0.213)	(0.088)				
Panel C: Labour						
	Population employed (%)	Cultivators (%)	Manual labourers (%)	Education (*ln*)	Manufacturing (*ln*)	Services (*ln*)
Irrigation	-1.391	-1.974	4.040	-0.086	0.082	0.066
(*standardised*)	(1.492)	(2.956)	(3.306)	(0.259)	(0.306)	(0.285)
Panel D: Demographics						
	Population density (*ln*)	Male share (%)	Child share (%)	SC share (%)		
Irrigation	0.303**	-0.337	1.121**	2.942		
(*standardised*)	(0.130)	(0.307)	(0.456)	(2.745)		
N	1868	1868	1868	1868	1868	1868

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on our key outcomes for villages with tube-wells built before 2001. Irrigation is measured as $litres/ha/day$ and standardised. Panel A reports results on the agricultural sector including: production in the Monsoon/$Kharif$ and Winter/*Rabi* season (EVI max log transformed), share of village area used for agriculture, whether the village grows water intensive and drought tolerant crops. Panel B reports results on consumption including: household asset ownership index and log of average night light luminosity. Panel C reports results on labour allocation, including: aggregate employment rate, share of the workforce in agriculture, and number of persons employed in village industries (log transformed). Panel D reports results on village demographics including: log of population density, population share of adults, children, and Scheduled Caste. The sample consists of villages with tube-wells built prior to 2000 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Appendices

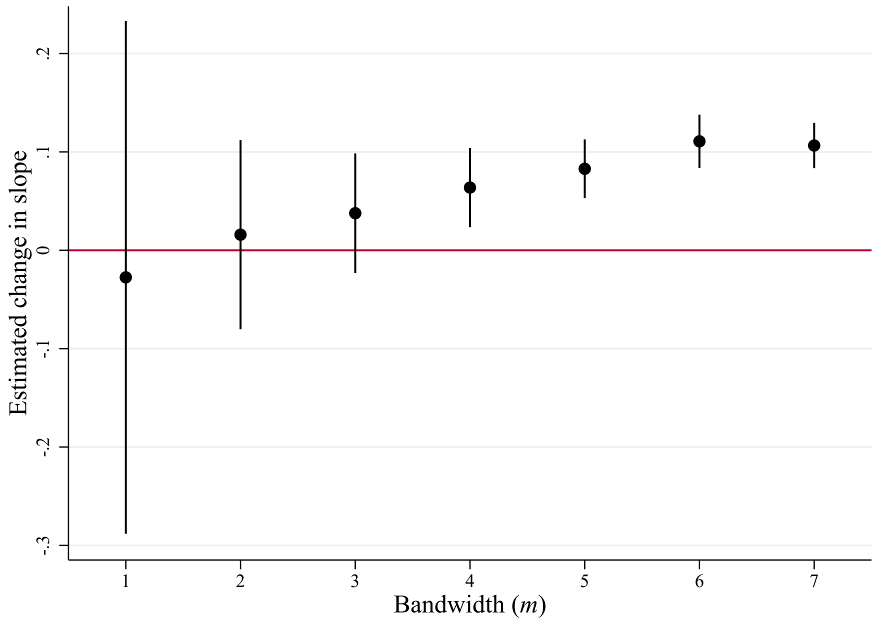
A. Additional Tables and Figures

Figure A1: Effect of efficiency and altitude on the maximum depth achievable by a centrifugal pump



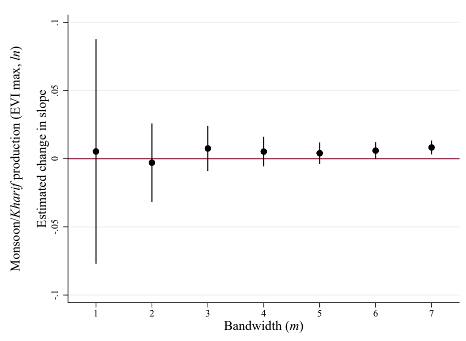
Notes: The maximum depth from which water can be extracted by a centrifugal pump can be calculated using Equation 2. 100% pump efficiency creates a perfect vacuum where P_2=0 $kg/m/s^2$. Atmospheric pressure is highest at sea-level where P_1=101,325 $kg/m/s^2$. Hence a pump working with 100% efficiency at sea-level achieves the maximum theoretical threshold from which water can be extracted – 10.33 metres. 75% and 50% pump efficiency indicate the corresponding percentage drop from a case of perfect vacuum. Altitude decreases atmospheric pressure (calculated using a barometric formula). The range of altitude plotted correspond to those found in the sample of villages.

Figure A2: Estimated kink in the deterministic relation between groundwater depth and irrigation at a range of bandwidths

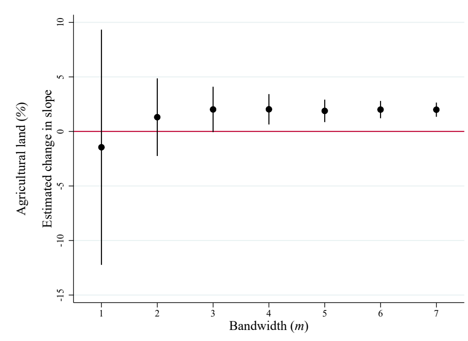


Notes: This figure presents point estimates and 90% confidence intervals on the effect of groundwater depth on irrigation at one meter interval bandwidths (Equation 3). Irrigation is calculated as water input in litres as an average over the year and standardised. Each regression includes state dummies and covariates (see Section 4 for details). Formal estimates of the kink for a bandwith of 7 metres are reported in Table 2. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample).

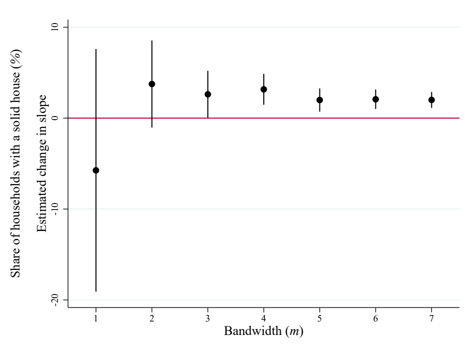
Figure A3: Estimated kink in the deterministic relation between groundwater depth and outcomes at a range of bandwidths



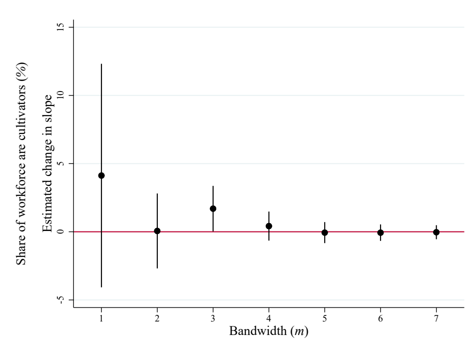
Panel A



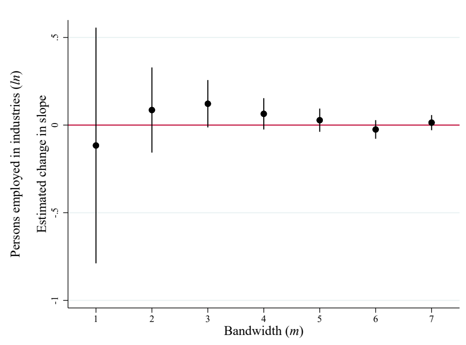
Panel B



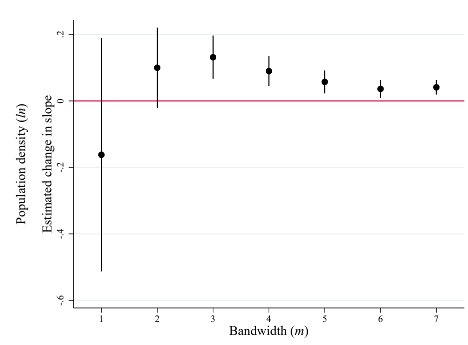
Panel C



Panel D



Panel E



Panel F

Notes: This figure presents point estimates and 90% confidence interval on the effect of groundwater depth on a selection of our outcome variables at one meter interval bandwidths (Equation 3). Each regression includes state dummies and covariates (see Section 4 for details). Formal estimates of the kink using the *fuzzy* RK specification with a bandwidth of 7 metres are reported in Tables 4, A6, 6, 7, and 9 for Panels A to F respectively. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample).

Table A1: Estimated kink in the deterministic relation of groundwater depth with pump adoption and irrigation when groundwater depth is measured over varying time horizons

	Pump adoption		Irrigation		
	Centrifugal pumps (*nb/ha*)	Submersible pumps (*nb/ha*)	Average (*standardised*)	Monsoon/ *Kharif* (*standardised*)	Winter/ *Rabi* (*standardised*)
	(1)	(2)	(3)	(4)	(5)
Panel A: Maximum groundwater depth in 2013					
δ_2	0.003*	-0.001	0.052***	0.058***	0.044***
	(0.002)	(0.002)	(0.013)	(0.013)	(0.013)
Panel B: Average maximum groundwater depth over 2010-2013					
δ_2	0.003**	0.001	0.107***	0.108***	0.102***
	(0.001)	(0.002)	(0.014)	(0.014)	(0.014)
Panel C: Average maximum groundwater depth over 2008-2013					
δ_2	0.004**	-0.001	0.096***	0.097***	0.092***
	(0.002)	(0.002)	(0.015)	(0.014)	(0.015)
N	3327	3327	3327	3327	3327

Notes: This table presents estimates on the effect of groundwater depth on pump adoption and irrigation when groundwater depth is measured over varying time horizons. δ_2 is the estimated change in slope of the assignment function at the kink point (Equation 3). Panel A presents estimates when the assignment variable is measured as the maximum groundwater depth recorded at any point in 2013. Panel B corresponds to the main specification reported in Table 2 when the assignment variable is measured as a three year average over 2010-2013. Panel C presents results when groundwater depth is measured as an average over five years (2008-2013). Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal (Column 1) and submersible (Column 2) pumps. Irrigation is measured as water input in litres and standardised. We report irrigation as an average over the year (Column 3), as well as independently for the Monsoon/*Kharif* (Column 4) and the dry Winter/*Rabi* season (Column 5). The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table A2: Estimated kink in the deterministic relation of groundwater depth with pump adoption and irrigation for different functional forms

	Pump adoption		Irrigation		
	Centrifugal pumps *(nb/ha)* (1)	Submersible pumps *(nb/ha)* (2)	Average *(standardised)* (3)	Monsoon/ *Kharif* *(standardised)* (4)	Winter/ *Rabi* *(standardised)* (5)
Panel A: Linear					
δ_2	0.003**	0.001	0.107***	0.108***	0.102***
	(0.001)	(0.002)	(0.014)	(0.014)	(0.014)
AIC	-6742.49	-5870.33	7927.19	7856.20	8044.51
R^2	0.36	0.21	0.37	0.39	0.35
Panel B: Quadratic					
δ_2	0.014***	0.001	0.009	0.020	-0.002
	(0.005)	(0.007)	(0.050)	(0.049)	(0.051)
AIC	-6743.96	-5866.36	7918.48	7849.16	8034.98
R^2	0.36	0.21	0.38	0.39	0.36
Panel C: Cubic					
δ_2	-0.007	-0.002	-0.255**	-0.241**	-0.265**
	(0.012)	(0.018)	(0.114)	(0.113)	(0.116)
AIC	-6743.96	-5866.36	7918.48	7849.16	8034.98
R^2	0.36	0.21	0.38	0.39	0.36
N	3327	3327	3327	3327	3327

Notes: This table presents estimates on the effect of groundwater depth on pump adoption and irrigation for varying functional forms. δ_2 is the estimated change in slope of the assignment function at the kink point (Equation 3). Panel A corresponds to the main specification reported in Table 2 when using a linear functional form. Panel B reports results when adopting a quadratic function and Panel C a cubic function. Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal (Column 1) and submersible (Column 2) pumps. Irrigation is measured as water input in litres and standardised. We report irrigation as an average over the year (Column 3), as well as independently for the Monsoon/*Kharif* (Column 4) and the dry Winter/*Rabi* season (Column 5). The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). AIC is the Akaike Information Criterion. Heteroskedasticity robust standard errors are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table A3: Impact of irrigation on agricultural production with varying proxies

	EVI max 2013 (*ln*) (1)	EVI max 2013 (*level*) (2)	EVI Differenced 2013 (*ln*) (3)	EVI max 2012-14 (*ln*) (4)	EVI max 2011-15 (*ln*) (5)
Panel A: Monsoon/*Kharif* season					
Irrigation	0.077***	383.953***	0.173**	0.053*	0.038
(*standardised*)	(0.030)	(144.906)	(0.081)	(0.030)	(0.028)
Mean	4609.952	4609.952	2361.142	4548.341	4538.157
SD	949.797	949.797	1084.559	902.152	875.150
Panel B: Winter/*Rabi* season					
Irrigation	-0.005	-60.399	-0.017	0.005	0.005
(*standardised*)	(0.026)	(152.321)	(0.037)	(0.026)	(0.027)
Mean	4609.952	4609.952	2361.142	4548.341	4538.157
SD	949.797	949.797	1084.559	902.152	875.150
N	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on agricultural production for varying specifications. Irrigation is calculated as *litres/ha/day* and standardised. Panel A reports results for the Monsoon/*Kharif* season, and Panel B for the dry Winter/*Rabi* season. Column 1 corresponds to the main specification reported in Table 4 where EVI is calculated as the maximum value (log transformed) in 2013. Column 2 reports results on the maximum EVI value in level form. Column 3 uses an alternative EVI proxy which is the difference (log transformed) between early season and the maximum in 2013. Column 4 measures the EVI maximum value over a 3 year period and Column 5 over a 5 year period. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation are reported for the full sample on the level form of the variables. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A4: Impact of irrigation on the distribution of landholdings

	Landless (%) (1)	0-2 Acres (%) (2)	2-4 Acres (%) (3)	4+ Acres (%) (4)
Irrigation	-5.121	4.887	-0.725	0.959
(*standardised*)	(4.436)	(3.329)	(1.530)	(2.508)
Mean	56.436	20.994	9.599	12.970
SD	22.177	17.540	7.725	13.449
N	2349	2349	2349	2349

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on the distribution of landholdings. Irrigation is measured as $litres/ha/day$ and standardised. Results are reported for four categories of land acreage in Columns 1 to 4 respectively – 0, 0-2, 2-4, and over 4. Each variable is calculated as the percentage share of households who own that specific landholding size. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A5: Impact of irrigation on village amenities

	Primary school (*binary*) (1)	Hospital (*binary*) (2)	Commercial bank (*binary*) (3)	Paved road (*binary*) (4)	Market (*binary*) (5)
Irrigation	-0.104	-0.041	-0.028	0.070	-0.012
(*standardised*)	(0.081)	(0.034)	(0.052)	(0.076)	(0.074)
Mean	0.949	0.044	0.868	0.276	0.264
SD	0.220	0.204	0.339	0.447	0.441
N	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on village amenities. Irrigation is measured as $litres/ha/day$ and standardised. We consider the availability of five key services captured by binary variables, reported in Columns 1 to 5 respectively, for whether a village has a: primary school, hospital, paved road, commercial bank, and regular market. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A6: Impact of irrigation on ownership of household assets

	Solid house (%) (1)	Refrigerator (%) (2)	Vehicle (%) (3)	Phone (%) (4)
Irrigation	19.772***	-0.209	1.728	0.948
(*standardised*)	(5.919)	(2.246)	(3.027)	(3.772)
Mean	44.444	8.717	21.288	72.581
SD	29.047	13.027	16.160	22.451
N	2349	2349	2349	2349

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on ownership of assets. Irrigation is measured as $litres/ha/day$ and standardised. Results are reported for four assets in Columns 1 to 4 respectively – solid house, refrigerator, vehicle, and phone. Each variable is calculated as the percentage share of households who own that specific asset. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A7: Impact of irrigation on the spatial distribution of agricultural sector employment by varying distance of nearest neighbour

	Nearest neighbour within 2km		Nearest neighbour within 5km		Nearest neighbour within 10km	
	Cultivators (1)	Labourers (2)	Cultivators (3)	Labourers (4)	Cultivators (5)	Labourers (6)
Panel A: Share of population employed (%)						
Irrigation	1.935	3.796	-1.605	1.897	-3.450	3.435
(*standardised*)	(4.808)	(5.843)	(2.481)	(2.830)	(2.225)	(2.554)
Mean	11.186	15.325	13.573	17.267	14.019	17.538
SD	12.177	14.423	13.027	15.007	13.215	15.015
Panel B: Share of workforce (%)						
Irrigation	7.982	8.179	-2.928	4.871	-5.475	7.519
(*standardised*)	(9.246)	(10.996)	(4.767)	(5.097)	(4.245)	(4.576)
Mean	26.283	35.229	29.893	37.065	30.468	37.210
SD	24.596	28.261	25.020	27.328	25.080	27.175
Panel C: Share of full-time workers (%)						
Irrigation	20.590	15.333	4.263	17.910**	3.753	16.540**
(*standardised*)	(13.714)	(15.206)	(6.550)	(7.553)	(5.754)	(6.646)
Mean	69.712	46.877	73.839	53.745	74.823	54.572
SD	37.270	38.719	35.235	38.062	34.756	37.866
N	757	757	2287	2287	2800	2800

Notes: This table presents *fuzzy* RK estimates on the spatial distribution effect of irrigation on agricultural sector employment. These effects are measured for the nearest neighbouring village without access to groundwater irrigation. Irrigation is measured as $litres/ha/day$ and standardised. Panel A reports results on the percentage share of the population employed, calculated as the ratio of those employed to the total working age population. Panel B reports results on the percentage share of the workforce, calculated as the ratio of those employed to the total workforce. Panel C reports results on the percentage share of full-time workers (those that work for more than 6 months of the year), calculated as the ratio of full-time workers to the total workforce. We consider two specific occupational categories in agriculture: cultivators are those who cultivate their own land, and labourers are those who work for a daily wage. All variables refer to total persons. The sample consists of villages without tube-wells in 2013 that are the nearest neighbour within 2km (Columns 1 to 2), 5km (Columns 3 to 4), and 10km (Columns 5 to 6) distance from our main sample of villages (villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point, see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the nearest neighbour sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A8: Impact of irrigation on the spatial distribution of agricultural production

	Nearest neighbour within 2km		Nearest neighbour within 5km		Nearest neighbour within 10km	
	Monsoon/ *Kharif* (EVI max, *ln*)	Winter/ *Rabi* (EVI max, *ln*)	Monsoon/ *Kharif* (EVI max, *ln*)	Winter/ *Rabi* (EVI max, *ln*)	Monsoon/ *Kharif* (EVI max, *ln*)	Winter/ *Rabi* (EVI max, *ln*)
	(1)	(2)	(3)	(4)	(5)	(6)
Irrigation	0.076	-0.101	0.032	-0.006	0.045	0.000
(*standardised*)	(0.063)	(0.069)	(0.034)	(0.031)	(0.032)	(0.029)
Mean	4836.310	4745.177	4674.697	4854.239	4630.789	4852.024
SD	860.850	964.735	929.288	1005.832	948.541	1031.849
N	757	757	2287	2287	2800	2800

Notes: This table presents *fuzzy* RK estimates on the spatial distribution effect of irrigation on agricultural production. These effects are measured for the nearest neighbouring village without access to groundwater irrigation. Irrigation is measured as $litres/ha/day$ and standardised. Using EVI, an index of vegetation cover from satellite imagery, we proxy for agricultural production by taking the maximum value of the index (log transformed) in both the Monsoon/*Kharif* and the dry Winter/*Rabi* season of 2013. The sample consists of villages without tube-wells in 2013 that are the nearest neighbour within 2km (Columns 1 and 2), 5km (Columns 3 and 4), and 10km (Columns 5 and 6) distance from our main sample of villages (villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point, see Section 3 for details on variable construction and sample). Mean and standard deviation reported on the level form of the variables for the nearest neighbour sample. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A9: Impact of irrigation on the village age distribution

	1-10 (1)	11-20 (2)	21-30 (3)	31-40 (4)	41-50 (5)	51-60 (6)
Panel A: Population (*ln*)						
Irrigation	0.481**	0.466**	0.441*	0.402*	0.417*	0.420*
(*standardised*)	0.235	0.234	0.231	0.228	0.227	0.225
Mean	745.062	838.726	732.307	587.786	455.378	297.289
SD	791.092	863.973	749.017	604.844	478.806	327.405
Panel B: Share of the population (%)						
Irrigation	0.007	0.004	0.000	-0.007	-0.002	-0.002
(*standardised*)	0.007	0.006	0.005	0.004	0.005	0.004
Mean	0.204	0.229	0.200	0.160	0.125	0.082
SD	0.046	0.033	0.024	0.021	0.026	0.022
N	2349	2349	2349	2349	2349	2349

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on the village age distribution. The disaggregated information on age is obtained from micro data collected by the Socio Economic Caste Census of 2012. Irrigation is measured as $litres/ha/day$ and standardised. Panel A presents results on the number of persons (log transformed). Panel B reports estimates on the share of population. We consider six 10-year age brackets reported in Columns 1 to 6 respectively. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 m) of the kink point (see Section 3 for details on variable construction and sample). Mean and standard deviation reported for the full sample, and in the case of population on the level form of the variables. The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A10: Impact of irrigation on the village economy; excluding covariates

Panel A: Agriculture	Monsoon/*Kharif* production (EVI max, *ln*)	Winter/*Rabi* production (EVI max, *ln*)	Agricultural land (%)	Water intensive (*binary*)	Drought tolerant (*binary*)	
Irrigation (*standardised*)	0.085***	0.003	17.590***	0.132*	-0.197**	
	(0.030)	(0.026)	(4.474)	(0.074)	(0.080)	
Panel B: Consumption	Household assets (*index*)	Night light (*ln*)				
Irrigation (*standardised*)	0.382*	0.090				
	(0.207)	(0.083)				
Panel C: Labour	Population employed (%)	Cultivators (%)	Manual labourers (%)	Education (*ln*)	Manufacturing (*ln*)	Services (*ln*)
Irrigation (*standardised*)	-2.358*	-0.350	2.138	-0.069	0.489*	0.388
	(1.399)	(2.906)	(3.183)	(0.231)	(0.293)	(0.270)
Panel D: Demographics	Population density (*ln*)	Male share (%)	Child share (%)	SC share (%)		
Irrigation (*standardised*)	0.403***	-0.252	0.749*	7.792***		
	(0.128)	(0.282)	(0.442)	(2.791)		
N	3327	3327	3327	3327	3327	3327

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on our key outcomes when excluding covariates from the specification. Irrigation is measured as *litres/ha/day* and standardised. Panel A reports results on the agricultural sector including: production in the Monsoon/*Kharif* and Winter/*Rabi* season (EVI max log transformed), share of village area used for agriculture, whether the village grows water intensive and drought tolerant crops. Panel B reports results on consumption including: household asset ownership index and log of average night light luminosity. Panel C reports results on labour allocation, including: aggregate employment rate, share of the workforce in agriculture, and number of persons employed in village industries (log transformed). Panel D reports results on village demographics including: log of population density, population share of adults, children and Scheduled Castes. The sample consists of villages with tube-wells in 2013 and groundwater depth within the bandwidth (7 *m*) of the kink point (see Section 3 for details on variable construction and sample). The specification includes only state dummies and *not* covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A11: Impact of irrigation on the village economy; excluding outliers

Panel A: Agriculture

	Monsoon/*Kharif* production (EVI max, *ln*)	Winter/*Rabi* production (EVI max, *ln*)	Agricultural land (%)	Water intensive (*binary*)	Drought tolerant (*binary*)
Irrigation	0.085***	-0.014	20.638***	0.125*	-0.196**
(*standardised*)	(0.030)	(0.026)	(4.759)	(0.074)	(0.078)

Panel B: Consumption

	Household assets (*index*)	Night light (*ln*)
Irrigation	0.598***	0.136
(*standardised*)	(0.212)	(0.084)

Panel C: Labour

	Population employed (%)	Cultivators (%)	Manual labourers (%)	Education (*ln*)	Manufacturing (*ln*)	Services (*ln*)
Irrigation	-2.619*	-0.854	1.095	-0.050	0.597**	0.403
(*standardised*)	(1.451)	(2.933)	(3.189)	(0.237)	(0.303)	(0.277)

Panel D: Demographics

	Population density (*ln*)	Male share (%)	Child share (%)	SC share (%)		
Irrigation	0.442***	-0.274	0.964**	7.479***		
(*standardised*)	(0.130)	(0.289)	(0.453)	(2.853)		
N	3092	3092	3092	3092	3092	3092

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on our key outcomes when excluding outliers. Outliers are captured using fluctuation in the maximum groundwater depth over a decade (2000-2010). We remove villages in the bottom 10 percentile of the distribution (corresponding to a drop in groundwater depth by more than 4 metres between 2000 and 2010). Irrigation is measured as $litres/ha/day$ and standardised. Panel A reports results on the agricultural sector including: production in the Monsoon/$Kharif$ and Winter/*Rabi* season (EVI max log transformed), share of village area used for agriculture, whether the village grows water intensive and drought tolerant crops. Panel B reports results on consumption including: household asset ownership index and log of average night light luminosity. Panel C reports results on labour allocation, including: aggregate employment rate, share of the workforce in agriculture, and number of persons employed in village industries (log transformed). Panel D reports results on village demographics including: log of population density, population share of adults, children and Scheduled Castes. The sample consists of villages with tube-wells in 2013, groundwater depth within the bandwidth (7 m) of the kink point, and a maximum 4 metre drop in the decadal groundwater depth (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table A12: Impact of irrigation on the village economy; excluding high altitudes

Panel A: Agriculture	Monsoon/*Kharif* production (EVI max, *ln*)	Winter/*Rabi* production (EVI max, *ln*)	Agricultural land (%)	Water intensive (*binary*)	Drought tolerant (*binary*)	
Irrigation	0.073**	-0.001	21.912***	0.175**	-0.208**	
(*standardised*)	(0.032)	(0.028)	(5.170)	(0.080)	(0.084)	
Panel B: Consumption	Household assets (*index*)	Night light (*ln*)				
Irrigation	0.480**	0.119				
(*standardised*)	(0.223)	(0.092)				
Panel C: Labour	Population employed (%)	Cultivators (%)	Manual labourers (%)	Education (*ln*)	Manufacturing (*ln*)	Services (*ln*)
Irrigation	-2.353	-0.325	2.415	0.038	0.601*	0.561*
(*standardised*)	(1.492)	(3.093)	(3.373)	(0.254)	(0.324)	(0.298)
Panel D: Demographics	Population density (*ln*)	Male share (%)	Child share (%)	SC share (%)		
Irrigation	0.433***	-0.158	0.749	7.944***		
(*standardised*)	(0.137)	(0.298)	(0.461)	(2.907)		
N	2999	2999	2999	2999	2999	2999

Notes: This table presents *fuzzy* RK estimates on the effect of irrigation on our key outcomes when excluding villages at high altitudes. We remove villages in the top 10 percentile of the altitude distribution (corresponding to altitudes above 600 metres). Irrigation is measured as *litres/ha/day* and standardised. Panel A reports results on the agricultural sector including: production in the Monsoon/*Kharif* and Winter/*Rabi* season (EVI max log transformed), share of village area used for agriculture, whether the village grows water intensive and drought tolerant crops. Panel B reports results on consumption including: household asset ownership index and log of average night light luminosity. Panel C reports results on labour allocation, including: aggregate employment rate, share of the workforce in agriculture, and number of persons employed in village industries (log transformed). Panel D reports results on village demographics including: log of population density, population share of adults, children and Scheduled Castes. The sample consists of villages with tube-wells in 2013, groundwater depth within the bandwidth (7 *m*) of the kink point, and less than 600 metres in altitude (see Section 3 for details on variable construction and sample). The specification includes state dummies and covariates (see Section 4 for details). Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

B. Decision Making Framework

In this Appendix we introduce a simple decision making framework for the adoption of different irrigation pumping technologies available to farmers. Consider a population of N farmers indexed by $i \in 1, ..., N$, living in a geographically diverse set of V villages indexed by $v \in 1, ..., V$. Each village has a given groundwater depth λ_v. While there are concerns of depleting aquifers in India, we show in Figure 1 that the annual average maximum groundwater depth among our sample of villages between 1996 to 2013 (which corresponds to the time period for which we also have records of tube-well construction) is stable around 7.5 metres.[36] Hence for ease of exposition we restrict the decision making to a one time choice when faced with a fixed groundwater depth. In this context, farmer i decides whether or not to invest in a single unit of irrigation when faced with his exogenous groundwater depth. We assume that one unit of irrigation is sufficient to irrigate the entire land endowment, l_i, of the farmer. Consequently, farmers with the most land get the highest returns from investment.

Based on Bernoulli's principle of fluid dynamics, we know from Equation 2 that there exists a maximum theoretical threshold from which water can be extracted with a centrifugal pump, k.[37] Deeper than this k threshold, no centrifugal pump can operate. Therefore if the water table depth in a given village exceeds k, the farmer must incur the cost r_s of a submersible pump if he chooses to irrigate. Conversely, when $\lambda_v \leq k$, a centrifugal pump will operate and thus enter the farmers' set of choices as a more cost effective technology since $r_c < r_s$. The functionality of a centrifugal pump however, will also depend on its' efficiency. This efficiency is random with known probability distribution $G(.)$ (and associated CDF $g(.)$) revealed to the farmer only at the time of purchase. Therefore, there exists a groundwater depth efficiency specific threshold, $e(\lambda)$, below which a centrifugal pump will not function. As such, there is a probability, $g(e(\lambda_v))$, that a farmer purchases a centrifugal pump which will not work.

When deciding on a technology, a farmer leverages all his current information. He also considers his forward looking expectations, including pump efficiency, relative costs, and yield increases from irrigation (which are assumed to be known to him). A risk neutral farmer will choose an irrigation technology simply to maximise his profits. In doing so, he compares the following profit functions – irrigating with a submersible pump ($\pi_{iv}^{I_s}(p, r_s, l_i)$), irrigating with a centrifugal pump ($\pi_{iv}^{I_c}(\lambda_v, p, r_c, l_i)$), or no irrigation ($\pi_{iv}^{N}(p, l_i)$) – which can

[36] Note that when including villages outside our sample bandwidth, we do observe a drop in the annual average maximum groundwater depth from 8 metres in 1996 to 10.5 metres in 2013.

[37] k corresponds to the difference in height outlined in Equation 2.

be written as:

$$
\begin{aligned}
\pi_{iv}^{I_s} &= pY_i^I l_i - r_s \\
\pi_{iv}^{I_c} &= (1 - g(e(\lambda_v)))(pY_i^I l_i - r_c) + g(e(\lambda_v))(pY_i^N l_i - r_c) \\
\pi_{iv}^{N} &= pY_i^N l_i
\end{aligned}
\tag{5}
$$

Where p, r_s, and r_c are the prices of output, a submersible pump, and a centrifugal pump respectively. Y_i^I denotes agricultural yields when irrigating and Y_i^N is for yields under no irrigation. As explained previously, a farmer is subject to a technology constraint such that $g(e(\lambda_v)) = 1$ if $\lambda_v > k$.

Given this framework, we consider three representative case scenarios: (1) a farmer whose liquidity constraint binds for both pump types, (2) a farmer who faces a liquidity constraint only for the more expensive submersible pump type, and (3) a farmer that is not liquidity constrained at all.

Case 1: Liquidity constrained for all irrigation technology. In this scenario, a farmer cannot access either irrigation technology. He therefore receives π_{iv}^N regardless of groundwater depth.

Case 2: Liquidity constrained for submersible pumps only. The farmer cannot afford the more expensive submersible pump. Therefore, if $\lambda_v > k$, he cannot access any irrigation technology. Alternatively, if $\lambda_v \leq k$, he will adopt a centrifugal pump when $\pi_{iv}^{I_c} > \pi_{iv}^N$. Expanding on these profit functions we show that:

$$
(1 - g(e(\lambda_v)))(pY_i^I l_i - r_c) + g(e(\lambda_v))(pY_i^N l_i - r_c) > pY_i^N l_i \tag{6}
$$

Rearranging Equation 6 demonstrates that a farmer will adopt a centrifugal pump if the increase in revenue with irrigation multiplied by the probability of the centrifugal pump working is larger than the cost of the pump:

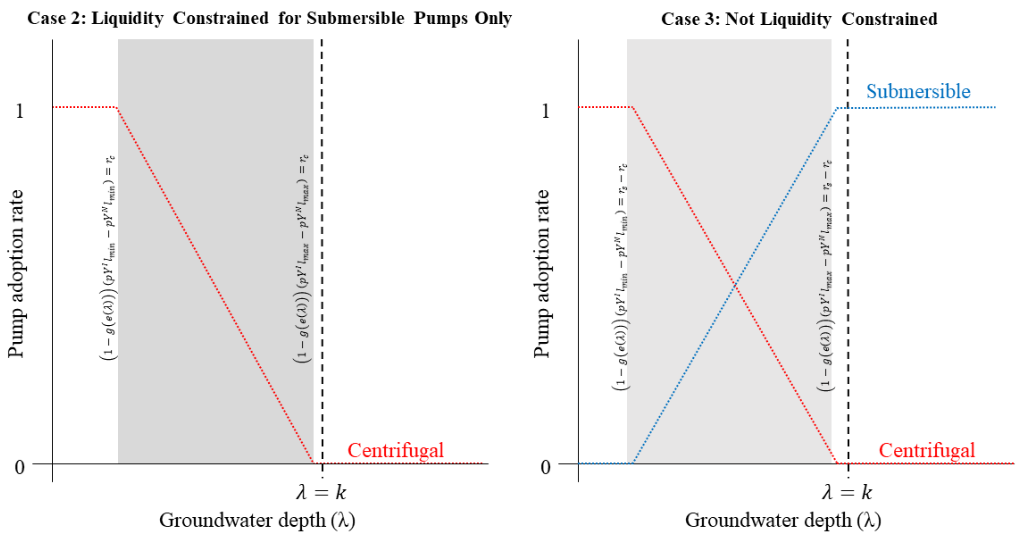
$$
(1 - g(e(\lambda_v)))(pY_i^I l_i - pY_i^N l_i) > r_c \tag{7}
$$

The probability of adoption therefore declines in $g(e(\lambda_v))$ up to the maximum theoretical threshold $\lambda_v = k$. Deeper than this threshold adoption is zero. Assuming $G(.)$ is uniformly distributed and the distribution of land holdings is orthogonal to λ_v, the decline in probability of adoption will be linear with a *kink* in the slope marginally before the maximum theoretical threshold.[38] Furthermore, as previously noted, given their higher

[38] Adoption will be zero when the largest farm is indifferent between adopting or not. That is, when:

marginal returns, farmers with the largest landholdings are most likely to adopt even when $\lambda_v - k$ is small.[39]

Figure B1: Illustrative diagram for the evolution of pump adoption with groundwater depth



Notes: This figure demonstrates the outcome of our decision making framework outlined in Section 2 for the adoption of different irrigation pumping technologies available to farmers (the smallest farmers are indexed by min and the largest by max). Y_i^I denotes agricultural yields when irrigating and Y_i^N is for yields under no irrigation. p, r_s, and r_c are the prices of output, a submersible pump, and a centrifugal pump respectively. λ_v is the exogenous village groundwater depth. k is the maximum theoretical threshold below which no centrifugal pump can operate. A farmer is subject to a technology constraint on centrifugal pumps such that $g(e(\lambda_v)) = 1$ if $\lambda_v > k$. It is the subset of farmers that are liquidity constrained for submersible pumps only (Case 2) that generate a decline in centrifugal pump adoption culminating in zero take-up at k.

Case 3: Not liquidity constrained. The farmer can purchase either of the irrigation technologies. If $\lambda_v > k$ a farmer will adopt a more expensive submersible pump when $\pi_{iv}^{I_s} > \pi_{iv}^N$ – that is, when the increase in revenue from irrigation is greater than the cost of a submersible pump. As a result adoption above the threshold is not dependent on groundwater depth:

$$(pY_i^I l_i - pY_i^N l_i) > r_s \tag{8}$$

$(1 - g(e(\lambda_v)))(pY_{max}^I l_{max} - pY_{max}^N l_{max}) = r_c$

[39] This corresponds closely with the report from the Minor Irrigation Census of 2013 which finds that the share of tube-wells owned by large farmers increases with the depth of the well (Rajan and Verma, 2017).

If $\lambda_v \leq k$ a farmer will adopt a submersible pump if $\pi_{iv}^{I_s} > \pi_{iv}^{I_c} > \pi_{iv}^{N}$. Therefore, a farmer who is not liquidity constrained and satisfies the condition in Equation 7 is now left to consider whether the certainty in submersible pump functionality justifies the difference in cost:

$$(1 - g(e(\lambda_v)))(pY_i^I l_i - pY_i^N l_i) > r_s - r_c \tag{9}$$

When the increase in revenue with irrigation multiplied by the probability of the centrifugal pump working is larger than the difference in cost between the two types of irrigation technology, a farmer will adopt the submersible pump. This condition leads to a substitution from centrifugal to submersible pumps as groundwater depth increases and the probability of the centrifugal pump working declines.

Figure B1 sketches how we expect adoption may evolve with groundwater depth within our decision making framework. Specifically, it is the subset of farmers that can afford a centrifugal pump but not a submersible (i.e. Case 2) that generates a decline in overall pump adoption culminating in zero take-up at the maximum theoretical threshold k. In the data we only observe what happens at aggregate when combining populations regardless of their liquidity constraints. However, given that the price of the cheapest submersible pump is half the average annual per capita consumption in our sample of villages, it is likely that liquidity will be a binding constraint for many farmers. We can therefore expect to observe a *kink* in the mapping of pump adoption and consequently irrigation with groundwater depth at the exogenous maximum theoretical threshold. We empirically demonstrate the presence and validity of this relationship in Section 4.

C. Data

C1 Irrigation

According to a standard engineering formula, three main factors affect water extraction from irrigation pumps – capacity, use, and well depth (Manring, 2013). We leverage data collected by the Fifth Minor Irrigation (MI) Census in 2013 on irrigation practices to calculate village-level indicators for pump capacity and usage.[40] Specifically, we measure pump capacity as the average horse power of pumps in a village. Usage is measured as the total number of pumping hours per day in a village.[41] We use our assignment variable – the maximum groundwater depth recorded at any point over a three year period (2010-2013) as our measure for well depth.[42] Using these three factors as outlined in Equation 10, we are able to calculate our main variable for irrigation in terms of water input in litres:

$$W_i(H_i D_i) = \rho \frac{P_i H_i}{D_i} \tag{10}$$

where i denotes a village, P_i is pump capacity, H_i is usage, and D_i is the depth from which water is lifted. The physical constant ρ, is given by:

$$\rho = c\frac{E}{dg} \tag{11}$$

where c is a constant to correct units and account for friction, E is pump efficiency, d is density of water, and g is the gravitational constant. Values for the constants used in the calculation of ρ are provided below in Table C1.

Calculated in this manner, we obtain a litres/day measure of groundwater extraction for irrigation. We then scale this by village size, generating a litres/ha/day variable. For the purpose of all our regressions, we further standardise this variable such that all results can be interpreted as the effect of a one standard deviation increase in irrigation.[43] To provide some context, one standard deviation is approximately equivalent to 103 $litres/ha/day$.

[40] Background information on each Census (e.g. questionnaires and instruction manuals on data collection) as well as official reports and aggregated statistical tables can be found on the official website of the MI Censuses at:http://micensus.gov.in. Village level data from each MI Census is publicly available in excel format on the Government of India open data platform at:http://data.gov.in

[41] Data on usage is available disaggregated by season. This allows us to calculate water input independently for both the monsoon/$Kharif$ and the winter/*Rabi* season. We obtain an annual measure by taking an average across the seasons.

[42] For information on how this data is compiled, refer to the part on groundwater in Section 3.

[43] To standardise the variable we subtract the mean and divide by the standard deviation of the sample for each observation.

Table C1: Constants used in water input calculation

Variable	Value	Units	Source
c	3.6×10^6		Ryan and Sudarshan (2022)
E	0.25		Ryan and Sudarshan (2022)
d	10^3	kg/m^2	Manring (2013)
g	9.81	m/s^2	Manring (2013)

Notes: The table shows the values of the constants used in the calculation of ρ in Equation 11. While density of water (d) and the gravitational constant (g) are standard in the literature (Manring, 2013), the values for pump efficiency (E) and friction (c) were obtained by Ryan and Sudarshan (2022) from case studies on irrigation pumping technology in India.

C2 Agricultural Production

Data on agricultural production based on direct field measurements is not available at the village level in India. We therefore rely on measures of vegetation cover calculated from satellite images as proxies for village agricultural yield. Specifically, we use data from the Enhanced Vegetation Index (EVI) estimated by the United States Geological Survey from images taken by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard NASA's Terra satellite. EVI appears to be the preferred index leveraged by most crop-mapping studies, as it accounts for atmospheric and background corrections (Gao et al., 2000). Evidence suggests that EVI values obtained from MODIS predict land use, in terms of classifying general crop types, with 90% accuracy (Wardlow and Egbert, 2010). Furthermore, evidence from Kouadio et al. (2014) indicates that EVI can successfully predict yield and demonstrates sensitivity to local variations in climate and geo-physical factors.

In order to determine the spatial distribution of plants from satellite images, the vegetation indices exploit the natural strong differences in plant reflectance. Specifically, the green photosynthetically active pigment in plant leaves – chlorophyll – strongly absorbs visible red light (RED). Conversely, the cell structure of leaves, strongly reflects near-infrared light (NIR). As a result, healthy vegetation absorbs most of the visible light that hits it and reflects a large portion of the near-infrared light. Therefore, in a given pixel, if there is more reflected radiation in the near-infrared wavelengths than in the visible wavelengths, we can concur that the vegetation cover is likely to be dense. For an in-depth review of the literature and methods on calculating vegetation indices based on satellite imagery, refer to Huete et al. (2002).

MODIS captures data in 36 spectral bands ranging in wavelength from 0.4 to 14.4 μm. The bands covering the wavelengths of interest for the purpose of capturing vegetation cover are generated at a global scale and a resolution of 250 m. Each image represents a 16-day composite, such that the value of each pixel is optimised following an algorithm which accounts for cloud cover obstruction, image quality, and viewing geometry. The images are published by the IRI/LDEO Climate Data Library.[44]

As part of their research evaluating India's national rural road expansion programme, Asher and Novosad (2020) compiled data on the EVI at the village-level for the years spanning 2000-2014. Specifically, the authors downloaded composite images for nine 16-day periods from June to October so as to cover the monsoon/*Kharif* growing season, and similarly from November to March so as to capture the winter/*Rabi* season. Each composite image was then spatially averaged to village boundaries. This data is made publicly available as part of the replication material of their published paper.[45] We leverage two proxies calculated from the index values in each growing season: (i) the maximum value (Labus et al., 2002), and (ii) the difference between the maximum value and the early season value (taken as the average of the first three 16 day periods) (Rasmussen, 1997). All proxies are measured in 2013, as well as an average over a three-year (2012-2014) and five-year (2011-2015) period. Finally, for more interpretable results, all proxies are log transformed.

As a validation test of these vegetation indices to proxy for agricultural production in the case of Indian villages, Asher and Novosad (2020) provide correlation estimates between the proxies and district level measures of agricultural output. Specifically, the authors ran panel regressions (2000-2006) of the EVI proxies on agricultural output obtained from the Planning Commission's district-wise domestic product data. An R-squared of over 70%, when using district-year fixed effects, suggests a strong correlation between the proxies and district level estimates of agricultural output.

C3 Consumption

Most developing countries do not collect detailed information on income or consumption as part of their censuses. As such, estimates of these economic indicators at a high geographic resolution are often unavailable at regular time intervals. Policy makers (especially the World Bank) and researchers have therefore recently relied on a method developed by Elbers et al. (2003) which uses an imputation rule derived from a household survey to generate small-area estimates of consumption in census data (Bedi et al., 2007). In a comparison of methods, McKenzie (2005) show that this prediction method through auxiliary surveys most accurately predicts non-durable consumption. Hentschel et al. (2000), demonstrate that this method produces unbiased estimates of poverty.

[44] Information on MODIS and images for Asia can be found on the site of the IRI/LDEO Climate Data Library:https://iridl.ldeo.columbia.edu/index.html?Set-Language=en

[45] The paper by Asher and Novosad (2020) and its associated dataset is available at:https://www.aeaweb.org/articles?id=10.1257/aer.20180268

Since the early 1990s the Government of India has conducted national socioeconomic censuses collecting information at both the individual and household level on caste, occupation, earnings, and assets, in order to determine the eligibility of households into various welfare schemes (Alkire and Seth, 2013). In 2012, the fourth such Socio Economic Caste Census (SECC) was implemented.[46] In that year, the India Human Development Survey-II (IHDS-II) was also conducted. It recorded direct measures of household consumption, as well as equivalent questions to the SECC on household assets and earnings.[47] Following the methodology of Elbers et al. (2003), Asher et al. (2021) use the IHDS-II data to predict household level consumption in the SECC dataset. Specifically, the researchers first estimate regressions of total household consumption on dummy variables of assets and earnings in the IHDS-II.[48] Coefficients from these regressions are then used to impute household level consumption values in the SECC. Finally, based on these household level values the researchers generate village level statistics for mean predicted consumption per capita and the share of the population below the poverty line.[49] Bootstrap estimates of these village level indicators are made available by the research team on the Socioeconomic High-resolution Rural-Urban Geographic (SHRUG, Version 1.5) open data platform for India.[50] We take these 1000 bootstrapped variables for predicted consumption per capita (for the purpose of the regression, these variables are log transformed) and share of the population below the poverty line, and run an additional bootstrap process on our main sample of villages when estimating the effect of access to irrigation on these indicators. As outlined in the work of Elbers et al. (2003), this bootstrapping process is required to obtain correct standard errors and p-values on our estimates.

Specific to our setting of Indian villages, Asher et al. (2021) provide three validation tests for the bootstrap estimates of consumption used in our analysis. First, the distribution of the consumption estimates at the village level matches broadly to that found in two national surveys conducted at the same time and at the same geographic level (IHDS-II and the National Sample Survey-2012). Second, there is a strong covariance between

[46]Information on the census can be found on the SECC website:`https://secc.gov.in/welcome`. Though the Government initially made the raw data public, only aggregated information is now available on the website.

[47]Information and data related to this survey can be found on the platform of Data Sharing for Demographic Research:`https://www.icpsr.umich.edu/web/pages/DSDR/index.html`

[48]These are the exact same variables as those recorded in the SECC. They include: type of roof and wall material, number of rooms, ownership of phone, house, vehicle, land, kisan credit card, and refrigerator, as well as the highest individual income in the household.

[49]The official poverty line for rural India is set at Rs.27/day, based on the Planning Commission's Tendulkar Committee Report in 2014.

[50]For detailed information on consumption data using the SHRUG open data platform, please refer to Asher et al. (2021). The dataset, including codebooks and references, can be found at:`http://www.devdatalab.org/shrug`

the district level predicted consumption estimates and those in the original household survey (IHDS-II). Third, by identifying how each component used in the imputation rule affects the difference in average consumption between the estimates and the original survey (IHDS-II), the researchers find that the transformation of asset ownership to consumption assumes a similar relationship across datasets. These findings provide confirmation that the predicted consumption estimates are valid proxies of the direct survey measures.

C4 Night Light

As an additional proxy for consumption, we leverage remote sensing imagery on Night-Time Light (NLT) at the village level across India. Initiated by the work of Henderson et al. (2011), NTL has since become a widely used proxy for economic activity. Researchers have adopted night-time luminosity to effectively capture GDP growth (Henderson et al., 2011), cross-sectional GDP (Bleakley and Lin, 2012), urbanisation (Harari, 2020), public expenditure (Hodler and Raschky, 2014), and employment (Mellander et al., 2015). In an analysis of Indian villages, Asher et al. (2021) find that night light is a highly statistically significant proxy for a range of development outcomes including - population, employment, per capita consumption, and electrification.

Night-time luminosity data is made available by the U.S. National Oceanographic and Atmospheric Administration (NOAA). The observations are assembled by the Operational Linescan System (OLS) aboard the Defense Meteorological Satellite Program (DMSP) satellites. A total luminosity value ranging from 0-63, is reported in grid cells covering a resolution of 1km x 1km. A description of the satellite instrumentation, data collection, and processing methods for NTL is detailed in the work of Elvidge et al. (1997). Asher et al. (2021) leverage this data to verify the effectiveness of night-time luminosity as a proxy for development indicators at the village level in India. As part of this work, the researchers compile a panel of NTL from 1994 to 2013 matched and aggregated to villages and towns across the country.[51] This dataset is made available by the research team on the Socioeconomic High-resolution Rural-Urban Geographic (SHRUG, Version 1.5) open data platform for India.[52] We make use of data on the maximum pixel luminosity at the village level. This value is captured for 2013 and is log transformed for ease of interpretation.

[51] The data is calibrated for consistent estimation across time, as suggested by Elvidge et al. (1997).

[52] For detailed information on NTL data using the SHRUG open data platform, please refer to Asher et al. (2021). The dataset, including codebooks and references, can be found at:`http://www.devdatalab.org/shrug`