Agricultural Productivity and Local Economic Development: Evidence from Private Investment in Irrigation

Camille Boudot-Reddy André Butler*

19th May 2022

Abstract

We study the average village level effects of access to groundwater for irrigation on the structural transformation of the rural economy. We exploit an absolute technological constraint imposed by the laws of physics on the operational capacity of pumps with depth of the water table in a *fuzzy* regression kink design. Our results show that increased access to irrigation – measured as an additional standard deviation unit ($\equiv 103\ litres/ha/day$) of groundwater – significantly boosts agricultural land production. Farmers respond to an increase in groundwater for irrigation by expanding their cultivated land area by 16%, and shifting away from drought tolerant crops. Furthermore, within the agricultural sector, we find that approximately 7 to 10% of cultivators and manual labourers respectively shift from part-time to full-time employment. This substantial shift in agricultural production translates into significant consumption gains including: (1) 0.45 standard deviation units increase in an index of durable assets, (2) 5% drop in the share of the village population living below the poverty line, and (3) increase in economic activity measured through night light. However, we find no evidence of labour re-allocation between different sectors of the local village economy.

JEL Codes: O10, O13, O53, Q15, Q25

*Boudot: University of Edinburgh (email: camille.boudot@ed.ac.uk); Butler: University of Edinburgh (email: andre.butler@ed.ac.uk). We thank Liang Bai, Irma Clots-Figueras, Maia Güell, Aprajit Mahajan, Ben Roth, and Andreas Steinhauer for helpful comments and conversations. We would also like to thank seminar participants at the University of Edinburgh, the Royal Economic Society Annual Conference (2022), the North East Universities Development Consortium Conference (Boston, 2021), the Econometrics Society Delhi Winter Meeting (2021), and the Scottish Economic Society Annual Conference (2021) for their feedback. This paper was previously circulated as"Watering the Seeds of the Rural Economy: Impact of Tube-Well Irrigation in India".

1. Introduction

A rich theoretical literature continues to debate the mechanism through which industrial growth is often preceded by unprecedented increases in agricultural output (Kuznets and Murphy, 1966; Johnston, 1970; Syrquin, 1988; Herrendorf et al., 2014). Various theoretical models suggest that in a closed economy, a boost to agricultural productivity generates demand for manufacturing goods which leads to a re-allocation of labour away from agriculture (Gollin et al., 2002; Kongsamut et al., 2001; Ngai and Pissarides, 2007). However this view has been challenged by research on open economies, which demonstrate that improved agricultural productivity can retard industrial growth as the economy specialises in the comparative advantage sector (Matsuyama, 1992; Field, 1978). Despite the centrality of this discussion in understanding the development process of low-middle income countries, there has been very little accompanying empirical evidence identifying the process by which shifts in agricultural productivity can alter the shape of the wider economy.

In this paper we provide direct empirical evidence on the effect of access to groundwater irrigation through private investment on agricultural production and the consequent structural transformation of the rural economy in India. Irrigation is one of the most conspicuous technologies for stimulating agricultural output. Primarily, improved productivity 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 income variability by reducing exposure to rainfall shocks (Sarsons, 2015), (2) augment land endowments by introducing the possibility of a second harvest (Blakeslee et al., 2020), and (3) complement other key inputs such as high yielding varieties (Gollin et al., 2021).

Groundwater contributes the largest supply of irrigation water globally (Siebert et al., 2010). In India, groundwater has been the main source of irrigation since the 1970s (Mukherji et al., 2013). Unlike dam and canal infrastructure which require government support for building and maintenance of the networks, groundwater can be accessed independently by farmers through private investment in tube-wells. Recent estimates suggest that groundwater now accounts for close to 70% of irrigated area across the country; compared to only 20% by canals (Jain et al., 2019). Exploitation of this resource however, is a growing concern – India has the fastest depleting aquifers in the world (Famiglietti, 2014; Rodell et al., 2018). An understanding of how access to groundwater irrigation affects the village economy, is therefore of crucial contemporaneous relevance.

In order to accurately ascertain the causal effect of irrigation on structural shifts in the rural economy we require an exogenous source of variation in access to groundwater. An absolute technological constraint imposed by the laws of physics on the operational capacity of centrifugal pumps with depth of the water table provides one such source of variation. There are two types of pump technologies available – centrifugal and submersible. Centrifugal pumps are installed on the surface and generate a pressure differential between the water table and the pumping mechanism. The maximum pressure differential achievable is created by a perfect vacuum in the pump mechanism. In this ideal state, Bernoulli's principle of fluid dynamics dictates that the maximum depth from which water can be extracted is a constant at any given atmospheric pressure (Faber, 1995). For instance, at sea level a pump generating a perfect vacuum will only be able to extract water from up to 10.33 meters below ground level. Extracting water from greater depths requires significantly more expensive submersible pumps,[1] which are placed at the bottom of a tube-well and pushes the water to the surface.

If all centrifugal pumps were homogenous in their ability to generate a perfect vacuum, we would expect to see a sharp discontinuity in access to irrigation at a specific groundwater depth with altitude. However, not all pumps operate at the same efficiency. For instance at sea level, the operational depth of a centrifugal pump falls from 10.33 to 5.18 meters as pump efficiency falls to half of its maximum potential. Consequently, there only exists a pump specific operational threshold; which we do not observe in the data. To guide our identification strategy, we therefore propose a simple decision making framework faced by farmers when deciding which irrigation technology to adopt. For groundwater depths shallower than 10.33 meters, the more cost effective centrifugal pumps are the farmers preferred choice. However as we approach the maximum theoretical depth for extracting water, fewer and fewer of these pumps will be operational. As a result, we predict a decline in centrifugal pump adoption with groundwater depth, culminating in zero adoption at the maximum theoretical threshold. Furthermore, under liquidity constraints, we predict an incomplete substitution to the more expensive submersible pumps. Given this trend in adoption of irrigation technology, we expect a kink in the mapping between groundwater depth and water extraction for agricultural irrigation to coincide with the maximum theoretical threshold. We exploit this quasi-random between-village variation in access to irrigation at an exogenous groundwater depth in a *fuzzy* regression kink design.

[1]Based on an online search, centrifugal pumps cost approximately INR 16,000 (GBP 160), half the price of submersible pumps.

Our empirical analysis makes use of both existing and newly-assembled datasets. Our assignment variable – groundwater level – is compiled using data published by the Central Ground Water Board (CGWB) of India which monitors wells four times a year across the country. We use the geographic positioning system (GPS) locations of these wells to match this data to multiple external contemporaneous datasets. Irrigation data, including ownership of centrifugal and submersible pumps, is compiled from the Minor Irrigation Census. These censuses form the longest spanning database of information on irrigation infrastructure at the village level. We draw from remote sensing, population and economic censuses, as well as administrative micro-data in order to describe village level agricultural production, amenities, demographics, sectoral labour allocation, as well as various measures of consumption. The benefit of a village-identified dataset is that it enables us to leverage spatial variation in groundwater at a high resolution and across a large geographic area. This is especially important in our empirical approach, as regression kink designs are notoriously demanding of sample size. However, a disadvantage of this data is that we are limited to broad categories of outcomes of interest and cannot disentangle our results on a finer scale.

We estimate the impact of access to irrigation as an additional standard deviation unit ($\equiv 103\ litres/ha/day$) of groundwater on our outcomes of interest. We find that increases in groundwater irrigation significantly boosts agricultural land production. Specifically, agricultural yield (derived from remote sensing vegetation indices) increases by 9.8% in the dry winter/*Rabi* growing season. Furthermore, farmers appear to re-optimise their production strategies by moving away from drought tolerant crops, as well as bringing an additional 16% of land area under cultivation. In accordance with other studies analysing the impact of irrigation on poverty, we find significant gains to consumption (Duflo and Pande, 2007; Sekhri, 2014). These are captured through three key variables: (1) share of the village population living below the poverty line drops by 5.4%, (2) durable asset index increases by 0.45 standard deviation units, and (3) average night-light intensity over a five-year period (a remote sensing proxy for economic activity) rises by 18%. Together, these results provide evidence for important returns to irrigation on consumption within the village. However in a heterogeneity analysis by size of landholding, we find that benefits from access to groundwater irrigation on agricultural production and consumption accrues mainly to villages with larger landholdings.

The impact of increased access to irrigation on the labour market includes a shift from part-time to full-time employment in the agricultural sector. Specifically, we find that groundwater irrigation increases the share of both cultivators and manual labourers working full-time by 7% and 10% respectively. An increase in agricultural productivity from access to irrigation, does not however appear to have strong effects on the sectoral transformation of the village economy. We find no evidence of increased employment, or number of village businesses established, across a range of sectors. Finally, we consider the potential effect of access to irrigation on migration. Villages with higher agricultural productivity may lead to a pooling in of labour, especially among working age men. While we find a significant increase in the village population, we do not detect any shifts when disaggregating the population by working age and gender.

Our paper is linked to a recent and growing literature providing empirical evidence on how productivity shocks in agriculture affect structural shifts in the economy. This literature builds on macro-economic models which demonstrate how agricultural productivity growth plays a key role in subsequent industrialisation, explaining cross-country disparities in income (Gollin et al., 2002, 2007; Córdoba and Ripoll, 2009; Vollrath, 2011). These results have been corroborated empirically by two quasi-experimental, cross-country studies on the impact of improved seed varieties (Gollin et al., 2021) and fertilizer use (McArthur and McCord, 2017). Specifically, Gollin et al. (2021) find that the impact of high yielding varieties on per capita GDP are associated with a rise in total factor productivity beyond those simply derived from a boost in crop yields, therefore partially attributing these to structural transformation. Analysing the effect of an increase in staple yields introduced by improved fertiliser use, McArthur and McCord (2017) show that this generated a minimum rise of 14% in GDP per capita and led to a 5% decline in the agriculture labour share over a five year period.

Evidence of the role that agriculture plays in structural transformation at a more micro level is mixed. In a study exploiting the spread of improved seed varieties in Brazil, Bustos et al. (2016) are able to isolate two different types of productivity shocks – labour-saving and land-augmenting. The authors find that adoption of labour-saving hybrid soy led to an expansion of employment in the local manufacturing sector. Conversely, land-augmenting maize led to an increase in the marginal product of labour in agriculture and consequently a reduction in industrial employment. Exploring the promotion of advanced wheat production technologies undertaken by Mussolini during his dictatorship, Carillo (2021) find significant long-run positive effects on industrialisation and economic prosperity that continue to persist today. In contrast, Foster and Rosenzweig (2004) estimate structural change from panel data on Indian villages, and show that boosts in agricultural productivity raise local wages and thereby hinder the potential for non-farm sectoral growth.

In two concurrent papers, Asher et al. (2021a) and Blakeslee et al. (2021) also aim to estimate the impact of access to irrigation on the structural transformation of the Indian rural economy. The identification in both of these papers relies on the unique geography of canal irrigation networks. As a result, they both identify the treatment effect of being in the command area of a canal. Blakeslee et al. (2021) find that structural transformation from access to canal irrigation depends largely on the presence of a nearby town. Specifically, they find that villages within a 2km distance of a town experience a drop in population density and employment in non-agricultural sectors, while villages further away experience a positive shift. Evaluating the long-run effects of canals, Asher et al. (2021a) find that structural change from agricultural productivity increases is associated with the formation and population growth of nearby towns. In contrast to these papers, our work focuses on the dominant source of irrigation – groundwater – which requires significant private investment from farmers. Consequently, we make use of more detailed census data on irrigation practices to capture the intensive margin; providing insight into the returns to private investment. Regardless of the type of irrigation, all three papers find very little evidence for within village labour reallocation while geographic movement of labour is dependent on access to outside opportunities.

Our paper is also closely linked to a strand of causally interpretable evidence on the impacts of access to irrigation. The scarcity of such research is due in large part to the empirical challenges involved in establishing reliable estimates. Our paper contributes to a limited literature using quasi-experimental methods to evaluate the impact of this key technology. Duflo and Pande (2007) analyse the distributional effects of irrigation dams in India – the authors found that while beneficiaries living downstream from irrigation dams increased their agricultural productivity and experienced lower levels of poverty, this was counter balanced by increases in poverty in upstream populations. Evidence from private investment in tube-wells for irrigation has been found to shift agricultural practices away from drought-tolerant crops increasing land productivity (Hornbeck and Keskin, 2014), while reducing water related conflict in the case of India (Sekhri, 2014). Furthermore, using randomly located geological formations that store pockets of water in the bedrock, Blakeslee et al. (2020) explore farmer adaptations to a drying up of these pockets of groundwater for irrigation. The authors found that in such cases, there is a consequential decline in farm income, however 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 describes the different technologies available to farmers for groundwater extraction, and outlines a simple decision making framework for the adoption of these technologies. 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 access to irrigation on the rural economy. Finally, Section 6 concludes.

2. Background

Private investment in groundwater irrigation, the focus of this paper, is the fastest growing source of irrigation water accounting for over 70% of the irrigated land in India. In this section, we first describe the technologies available to farmers for extracting groundwater. In particular, the role of groundwater depth in determining the most suitable pump for water extraction. We then outline a simple decision making framework faced by farmers when deciding which technology to adopt.

A. Irrigation Pumps

The technology most suitable for extracting groundwater depends on the depth of the water table in a given location. There are two main types of pump available, centrifugal and submersible. Centrifugal pumps are the most widely adopted due to their affordability. They are installed on the ground and create a vacuum with water moving up the tube from an area of high pressure at the bottom of the well, to an area of low pressure in the pumping mechanism (Figure 1). The extraction of water from a 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, under a perfect vacuum ($P_2 = 0\ kg/m/s^2$) and a given atmospheric pressure, the depth from which water can be extracted with a centrifugal pump is a constant. At sea level ($P_1 = 101,325\ kg/m/s^2$) this depth is 10.33 meters. In real world circumstances however, it is unlikely that all pumps are able to create a perfect vacuum. Specifically, as we demonstrate in Figure 2, the maximum pumping depth at sea level falls from 10.33 to 5.18 meters as pump efficiency falls to half its maximum potential. We can therefore expect that once the groundwater level falls below a pump efficiency specific threshold, 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. A submersible pump is placed at the bottom of the tube-well and pushes the water to the surface. Consequently, provided it has sufficient horsepower, a submersible pump can extract water from any depth. This functionality however, comes at more than twice the price of a centrifugal pump.[2]

B. Decision Making Framework

In this section we introduce a simple decision making framework for the adoption of different irrigation 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 level λ_v.[3] In this context, farmer i decides whether or not to invest in a single unit of irrigation when faced with his exogenous groundwater level. 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 the depth from which water can be extracted is a constant. At sea level, this depth is 10.33 meters. Below this threshold, k, no centrifugal pump can operate.[4] 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 $\lambda_v < k$, a centrifugal pump will operate and thus enter the farmers' set of choices as a more cost effective technology; $r_c < r_s$. However, the

[2]Based on an online search for a range of pump models for the Indian market, we found that the entry price for a centrifugal pump is approximately INR 16,000 (equivalent to GBP 160), while a submersible pump costs over INR 30,000 (GBP 300). To put these costs into context – the mean annual per capita consumption in our sample of villages is approximately INR 18,000 (GBP 180).

[3]In reality the water table fluctuates temporally, however for ease of exposition we will just consider a one time choice when faced with a fixed groundwater level.

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

functionality of a centrifugal pump will 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. Additionally, there exists a groundwater specific efficiency 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). Specifically, 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 be written as:

$$\begin{aligned} \pi_{iv}^{I_s} &= pY_i^I l_i - r_s \\ \pi_{iv}^{I_c} &= (1 - g(e(\lambda_v)))(pY_i^I l_i - r_c) + g(e(\lambda_v))(pY_i^N l_i - r_c) \\ \pi_{iv}^{N} &= pY_i^N l_i \end{aligned} \tag{3}$$

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:

$$(1 - g(e(\lambda_v)))(pY_i^I l_i - r_c) + g(e(\lambda_v))(pY_i^N l_i - r_c) > pY_i^N l_i \tag{4}$$

Rearranging Equation 4 demonstrates that a farmer will adopt a centrifugal pump if the increase in revenue with irrigation multiplied by the probability of the pump working is larger than the cost of the pump:

$$(1 - g(e(\lambda_v)))(pY_i^I l_i - pY_i^N l_i) > r_c \tag{5}$$

The probability of adoption therefore declines in $g(e(\lambda_v))$, up to the threshold $\lambda_v = k$. Above 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 threshold.[5] Furthermore, as previously noted, given their higher marginal returns, farmers with the largest landholdings are most likely to adopt even when $\lambda_v - k$ is small.

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:

$$(pY_i^I l_i - pY_i^N l_i) > r_s \tag{6}$$

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 5, is now left to consider whether the certainty in submersible pump functionality justifies the difference in cost:

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

Similar to Case 2 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

[5] Adoption will be zero when the largest farm is indifferent between adopting or not. That is, when: $(1 - g(e(\lambda_v)))(pY_{max}^I l_{max} - pY_{max}^N l_{max}) = r_c$

increases and the probability of the centrifugal pump working declines.

Figure 3 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 kink in overall pump adoption and consequently irrigation. Additionally, under the assumption that when $g(e(\lambda_v)) = 0$ – that is, centrifugal pumps work with certainty – then Equations 5 and 6 hold even for the smallest farmers. On the other hand, as $\lambda_v - k$ becomes very small, only the farmers with the largest landholdings are most likely to adopt centrifugal pumps.[6]

In the data we only observe what happens at aggregate when combining populations regardless of their liquidity constraints. However, our model suggests that as long as a large proportion of the population are partially liquidity constrained (i.e. Case 2) then we can expect to observe a kink in pump adoption and consequently groundwater irrigation close to the threshold k. We empirically demonstrate the presence and validity of this relationship in Section 4. Our model also suggests that the returns to irrigation depends on landholding size, such that farmers with larger landholdings benefit more from private investment in groundwater irrigation. We test this prediction using a heterogeneity analysis by landholding size, reported in Section 5, and demonstrate that investing in irrigation has more meaningful implications for agricultural yield, and consequently consumption gains, in villages were the median landholding size is above average.

3. Data

For the purpose of this study, we have assembled a high resolution dataset including information on irrigation practices, as well as a range of features describing the rural economy. We link observational groundwater data from wells across the country with multiple external contemporaneous datasets to obtain a village level cross-section. Importantly for our empirical approach, this enables us to leverage spatial variation in groundwater at a high-resolution over a large geographical area.

Data on our assignment variable - groundwater level - come from the official website of the Central Ground Water Board (CGWB).[7] 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] In Figure 3, we index the decision equation for small farmers as min, and for large farmers as max.

[7] Data can be downloaded in excel format from: `http://cgwb.gov.in`. We accessed this data in June 2020.

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[8] – 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 covering 2011-2013.[9] As the water table fluctuates temporally, taking a three year horizon allows us to account for some of this variation. As a robustness test, we present results when constructing our assignment variable using a one and five year time period. Combining village boundary shapefiles offered by the Socioeconomic Data and Applications Center (SEDAC) of NASA,[10] along with the GPS coordinates of wells, we create a village level match. Specifically, we attribute the measure of our assignment variable to a village if the well falls within the village boundary.[11] Figure 4 presents a map of our final sample of matched wells across the country, as well as whether these fall below or above the operational threshold for centrifugal pumps. As Figure 4 plainly demonstrates, our data on groundwater level provide the basis of our empirical approach – evenly distributed high-resolution spatial variation across a large geographic coverage.

We compile data on irrigation practices from the Fifth Minor Irrigation (MI) Census conducted in 2013.[12] With the objective of collecting information to be used for the planning and management of water resources in the agricultural sector, the Government of India has implemented a MI Census every 7 years since 1986-87.[13] 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.[14] Specific to the needs of our study, the Fifth MI Census has data on ownership of different pump types, including submersible and centrifugal. Importantly, there also exists

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

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

[10]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.

[11]If more then one well was matched to the same village, an average of our assignment variable was taken.

[12]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. We accessed this data in June 2020.

[13]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.

[14]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 largely include dam and canal irrigation projects.

information on pump capacity (horse power) and usage (pumping hours), which allow us to calculate water input in litres following a standard engineering formula (Manring, 2013) (see Appendix B for detail on the construction of this variable). This measure of irrigation is also used in the recent work of Ryan and Sudarshan (2020) evaluating the effect of groundwater rationing in Rajasthan.

Data on agricultural production based on direct field measurements is - to the best of our knowledge - not available at the village level in India. We therefore rely on measures of vegetation cover calculated from satellite images as a proxy for agricultural yield at the village. Specifically, we use data from the Normalised Difference Vegetation Index (NDVI) and 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 in the recent work by Asher and Novosad (2020) for their evaluation of India's national rural road expansion programme, and was made available by the authors as part of their replication dataset.[15] The authors extracted information on NDVI and 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 matched to village boundary shapefiles. We leverage this raw data to calculate three proxies for each vegetation index in each season – mean, maximum, and the difference between early-season (taken as the mean of the first three 16-day periods) and the maximum value observed. By differencing out non-crop vegetation (such as forest cover) this latter proxy provides the most reliable measure of crop production. Appendix B includes further information on how the indices and proxies are constructed, a discussion of the literature on using remote sensing imagery to predict crop yields, 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 access, 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

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

in each village.[16] We use this information to create three binary measures of crop choice: does a village grow winter crops, drought tolerant crops, or cash crops.[17] In terms of agricultural inputs, we also draw upon the 2011 Village Directory for our measures 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 (SECC) of India in 2012. 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).[18]

We draw on the 2011 Population Census for information on labour allocation of village residents.[19] Specifically, we obtain data on total employment, as well as for two occupational categories of employment in the agricultural sector – cultivators and manual labourers. Cultivators are those that cultivate their own land, while manual 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. Specifically, 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 businesses at the village level, we make use of data from the Sixth Economic Census conducted in 2013.[20] The Economic Census is the only complete enumeration of all economic establishments in India, formal and informal, with

[16]Data from the 2011 Population Census Village Directory can be downloaded from:https://censusindia.gov.in/2011census/censusdata2k11.aspx. We accessed this data in June 2020.

[17]Winter season crops include: wheat, barley, potato, oilseed, and chickpea. Drought tolerant crops (based on classification by the International Crops Research Institute for Semi-Arid Tropics) include: millet, sorghum, maize, pigeon pea, and groundnut. Cash crops (these cannot be directly used for household consumption as they require post-harvest processing, but are generally considered to be more profitable) include: sugarcane, oilseed, cotton, and tobacco.

[18]As mentioned previously, the paper by Asher and Novosad (2020) evaluating India's national rural road construction programme and its associated dataset is available at:https://www.aeaweb.org/articles?id=10.1257/aer.20180268

[19]Data from the 2011 Population Census can be downloaded at: https://censusindia.gov.in/2011-Common/CensusData2011.html. We accessed this data in June 2020.

[20]This data is available on the National Data Archive site: http://microdata.gov.in/nada43/index.php/catalog/47. We accessed this data in June 2020.

no restrictions on size or location.[21] Detailed records are kept on employment and business characteristics (such as industry classification), but not on wages, inputs, or outputs. Compared to the Population Census which includes employment records for village residents even if these take place outside the village, the Economic Census concentrates on economic activity only in village businesses. We calculate the share of the workforce employed in the following sectors: agro-processing (this excludes crop production), livestock, construction, manufacturing, and services. Among our sample, these five industries account for over 70% of those employed in village businesses. Additionally, we also report our results on the total number of economic establishments in these industries.

In order to capture shifts in village demographics from migration in response to access to groundwater irrigation, we consider four population indicators. We leverage the Population Census of 2011 to obtain data on the total village population and to calculate the population density. Additionally, we make use of the disaggregated population data by age and gender, recorded by the SECC 2012, to capture the share of the total population as well as the male population which are of working age (15 to 65 years).

As a final set of outcomes of interest, we look at a range of consumption indicators. These are all obtained from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG, Version 1.5).[22] Night light, measured 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) (see Appendix B for a detailed discussion of this variable). For a more direct measure, we leverage household micro-data collected by the India Human Development Survey-II in 2012 to predict consumption on a range of asset and income variables equivalent to those recorded in the SECC 2012 (these predicted variables are available in the SHRUG). Following the methodology suggested by Elbers et al. (2003), we can then impute consumption using the SECC asset and income data to generate village-level statistics for predicted consumption per capita and the poverty rate (refer to Appendix B for further detail on how these indicators are generated and a discussion of the literature on the methodology). Finally, we look at an index of asset ownership – as recorded by the SECC 2012 – as well as each individual major asset independently.

[21] 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.

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

As a robustness test to our identification strategy, we show that there does not exist a kink in the relationship between groundwater level and a range of covariates. These covariates should capture natural endowments of the village unrelated to groundwater irrigation. Specifically, we consider: distance to the nearest river (obtained from the 2011 Village Directory), an index of maximum potential yield calculated using the agro-ecological zones' potential yield for 15 crops under rain-fed conditions (obtained from the SHRUG), and a binary indicator for whether or not the village has tube-wells in the command area of a dam and/or canal network (obtained from the Fifth MI Census).

We link all our different datasets at the village level. For variables coming from the Population Census or Economic Census, these are directly matched to the SHRUG Dataset of India (Version 1.5) using Census village identifiers. However for the wells and irrigation data, we use a combination of Python and Stata code for fuzzy matching on names adapted to the local Indian languages.[23] This method resulted in a matching success of approximately 80%. Table 1 provides summary statistics of all key variables on the final sample size of 4,896 villages across 477 districts for 19 States of India. These statistics suggest that in an average village, approximately 70% of the agricultural land is irrigated by tube-wells. Agriculture is on average the largest employer with approximately 13% of the workforce engaged as cultivators and 18% as manual labourers. Services on the other hand, despite being the largest employer among village businesses, engages only 8% of the workforce. Approximately 30% of the village population live below the poverty line.

4. Empirical Approach

In this paper, we are interested in capturing the effects of access to groundwater irrigation on the structural transformation of the rural economy. 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 irrigation. Any naive correlation estimates between irrigation and economic outcomes will in such a case be biased; partially attributing the effect of irrigation to markets rather than the technology itself. In this section we introduce our proposed empirical approach – *fuzzy* Regression Kink (RK) Design – to circumvent this endogeneity issue. Furthermore, we present graphical evidence and estimation results corroborating the validity of this method.

[23] 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

A. Regression Kink Design

To identify the true causal effect of irrigation, we use a *fuzzy* RK design which relies on quasi-experimental variation in the likelihood of having access to groundwater around a specific threshold.[24] In our empirical model, the assignment function determining the probability of being treated – that is having access to irrigation – is driven by a technological constraint in centrifugal pump capacity with groundwater depth. Specifically, 10.33 meters is the maximum theoretical groundwater depth – k – at which a centrifugal pump can extract water for irrigation at sea level.[25] Below this depth, the cost of irrigating increases significantly due to the price differential of the more expensive submersible pumps.[26] As described further in the next subsection; we demonstrate that for our sample of villages, there is in fact a discontinuity in the first-derivative of our assignment function at the given threshold.

In our analysis, the change in slope of the assignment function at the kink is unknown and must therefore be estimated based on observed data. Accordingly, we employ a *fuzzy* RK design, in which both the assignment variable and the treatment variable may be observed with error (Card et al., 2015b).[27] Specifically, we expect a kink in the deterministic relation between our treatment variable – irrigation, I – and our assignment variable – groundwater depth, W – at k. It follows that if irrigation exerts a causal effect on our outcome of interest – Y – we should expect to see an induced kink in the relationship between Y and W at k. Accordingly, the causal impact can be estimated by dividing the change in slope of the conditional expectation function for an outcome variable of interest (Equation 8), by the corresponding change in slope of the conditional expectation function for the assignment function (Equation 9) at the kink point. Specifically, we obtain the numerator of the *fuzzy* RK design estimand from the following parametric polynomial model:

[24] 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).

[25] k corresponds to the difference in height explained by Bernoulli's principle of fluid dynamics outlined in Equation 2.

[26] For detailed information on irrigation pump technologies available to farmers, including their costs, refer to Section 2.

[27] 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 with an estimate based on the observed data.

$$E[Y|W=w] = \mu_0 + \left[\sum_{p=1}^{\bar{p}} \gamma_p (w-k)^p + \nu_p (w-k)^p \cdot D \right] \quad (8)$$

where $D = \mathbb{1}[W \geq k]$ is a binary variable indicating whether the village experienced a groundwater level deeper than the threshold k at any point between 2011 to 2013. The change in slope of the conditional expectation function of Y at the kink point is given by ν_1.

Similarly, we estimate the denominator of the *fuzzy* RK design estimand using the following parametric polynomial form:

$$E[I|W=w] = \alpha_0 + \left[\sum_{p=1}^{\bar{p}} \omega_p (w-k)^p + \pi_p (w-k)^p \cdot D \right] \quad (9)$$

The change in slope of our treatment variable – I – at the kink point of the assignment variable W is captured by π_1. The impact of irrigation is therefore the ratio of the coefficients – $\beta = \nu_1/\pi_1$, and should be interpreted as the average treatment effect on the treated. Standard errors for β are recovered using the Delta method.

All our main regressions are estimated using a linear functional form ($\bar{p} = 1$). However as a robustness test, we compare our results when using a quadratic and cubic function. As suggested by Calonico et al. (2014), we do not include covariates as controls especially since these are shown not to be affected by the assignment function at the kink point. We do however, include state dummies in all regressions. Furthermore, all main regressions use a bandwidth (b) of 7 metres, such that $|w-k| \leq b$. Our results however, are shown to be robust to a range of bandwidth (down to 3 metres).

B. Impact of Groundwater Depth on Irrigation

Identification in a *fuzzy* RK design, requires two 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, and (2) the treatment assignment function is continuous at the kink point (i.e., there is no jump in the direct marginal effect of the assignment variable on the outcome of interest at the kink).[28]

Graphical evidence: We begin by showing graphical evidence to validate the *fuzzy*

[28] 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.

RK design assumptions. In response to the first assumption on the smooth density condition, we plot the probability density function of our assignment variable to check for manipulation of ones' position at the kink point. Note first that the exact location of the kink point is village specific as it varies with air pressure at different altitudes, as shown in Panel A of Figure 5.[29] Panel B of Figure 5 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 at the kink 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 the kink point.

The evolution of the relationship between centrifugal pump ownership and our assignment variable normalised at the kink point, provides evidence towards the second assumption. Corroborating the technological constraint faced by centrifugal pumps with groundwater depth, Panel A of Figure 6 demonstrates a clear kink in the slope of the relationship at the given threshold. As hypothesised in our decision making framework, 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 when groundwater depth exceeds the threshold ($w \geq k$). Furthermore, Panel B of Figure 6 shows a gradual increase in submersible pump ownership – likely driven by bigger wealthier farmers switching to this more expensive technology as the probability of a centrifugal pump operating declines. However, it is also clear that this technology substitution is incomplete. Specifically, centrifugal pump adoption (nb/ha) declines from 0.08 to 0.01, while submersible pump adoption only increases from 0.02 to 0.04 on the left of the threshold.

Importantly for our empirical approach, Panel C and D of Figure 6 exhibit a kink in the relationship between irrigation (measured as $litres/ha/day$) and our assignment variable. Specifically, as the depth of the water table increases, access to groundwater irrigation for agriculture declines, followed by a sharp switch to a constant level of groundwater extraction for depths greater than the threshold. This result provides confirmation that our empirical approach is capturing a shift in irrigation at a given threshold driven by an exogenous factor – the operational capacity of centrifugal pumps with groundwater

[29] Data on altitude was extracted from raster files for the whole of India, obtained from the ALOS Global Digital Surface Model. A barometric formula was used to calculate atmospheric pressure with varying altitude.

depth.[30]

We expect that if irrigation has a causal effect on our outcomes of interest, we should see an induced kink in the relationship between the outcome and groundwater depth at the threshold. Panel A of Figure 7 exhibits this graphical evidence for our measure of agricultural yield (NDVI-derived) in the dry winter/*Rabi* season. As the depth of the water table increases and access to groundwater irrigation declines, so does agricultural yields, followed by a sharp switch at the threshold. Conversely, with respect to covariates unrelated to irrigation, we expect that the conditional expectation of any such covariate should be twice continuously differentiable at the kink point. Panel D of Figure 7 exhibits this pattern for distance to the nearest river, a natural endowment of a village unrelated to the potential impact of irrigation.

Estimation results: Table 2 shows the results from our empirical specification outlined in Equation 9. We report the estimate for π_1, which corresponds to the change in slope of our treatment variable – irrigation from tubewells – at the kink point of the assignment variable – groundwater depth. Results based on a linear functional form suggests a statistically significant positive evolution of the relationship between groundwater depth and centrifugal pump adoption (Column 1) as well as groundwater extraction for irrigation at the kink point. This result is consistent when considering the average irrigation over the year (Column 3), as well as during the monsoon/$Kharif$ (Column 4) and the winter/*Rabi* season (Column 5) independently.

In support of our empirical strategy, we conduct various robustness tests. First, we demonstrate that for our covariates – distance to the nearest river, whether the village has pumps inside a canal command area, and maximum potential yield – there is no detectable change in the slope of the conditional expectation function at the kink point (see Columns 6 to 8 of Table 2). Second, we analyse the sensitivity of our results to the choice of polynomial order, reported in Panel B and C of Table A1. We find that the standard errors in the quadratic functional form increase substantially and the results in this form are no longer statistically significant. Third, we show that our results are consistent for our assignment variable calculated as the maximum groundwater depth recorded at any point over a one, three, or five year time horizon (see Table A2). Finally, we explore the

[30] As is explained in Appendix B, water extraction in litres (our measure of groundwater irrigation) is a function of pump capacity, pump usage and depth of the water table. Holding all other factors constant, as depth increases the flow rate of pumps will decrease. For depths deeper than the threshold, we observe a slight increase in the use of submersible pumps. However, this is likely to be offset by the increase in depth, thereby explaining the low correlation with pump adoption to the right of the threshold.

sensitivity of the deterministic relation between groundwater irrigation and depth of the water table at $w = k$ to the choice of bandwidth level. As shown in Figure A1, our results are consistent across bandwidth size down to 4 meters either side of the kink.

5. Results

In this section we report and discuss our results on the impact of groundwater irrigation on the structural transformation of the rural economy. 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 conditional expectation function at the kink point from Equation 9 and Equation 8 (see Section 4 for more detail on the estimation strategy). Our explanatory variable for access to irrigation is calculated as $litres/ha/day$ 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.

A. Agriculture

Before all else, we evaluate the impact of groundwater irrigation on agricultural yields. Leveraging two different vegetation cover indices – the Normalised Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) – we construct three alternative proxies of yield at the village level. These include the: mean, maximum, and difference between early-season and the maximum. This latter proxy is our preferred outcome variable for crop production. Previous research has shown that by differencing out non-crop vegetation such as forest cover, it provides a more direct estimate of agricultural yield (Rasmussen, 1997). Table 3 presents results on the effect of irrigation on these proxies for both the monsoon/$Kharif$ and the dry winter/*Rabi* growing season.

Focusing on the NDVI-derived differenced proxy for agricultural yield we find that irrigation has a positive impact on crop production. This effect is concentrated during the dry winter/*Rabi* season. We estimate an impact of 9.8% higher agricultural yield during the winter/*Rabi* season for a one standard deviation increase in groundwater irrigation use (Panel A Column 4 of Table 3). Graphical evidence presented in Panel A of Figure 7 corroborates this result. The plot demonstrates a sharp visible kink in the relationship between the winter/*Rabi* differenced NDVI-derived value for agricultural yield and normalised groundwater depth at the given threshold. This increase in yield however, is not detected by the EVI-derived proxies (Panel B of Table 3). This is possibly because EVI was partly developed to compensate for the effects of NDVI saturation over high biomass areas such as forests (Huete et al., 2002; Gao et al., 2000). As a result, EVI tends to present relatively lower ranges over lower biomass sites, making it less sensitive in semi-arid agricultural settings (see Appendix B for further details including a discussion of the literature on using remote sensing imagery to predict crop yields).

During the monsoon/$Kharif$ season, irrigation does not appear to have a significant impact on yield (when considering the differenced proxy, see Column 1 of Table 3). This is to be expected, since unlike the dry winter months the monsoon season receives heavy rainfall; hence reducing the reliance on irrigation for cultivation. Interestingly however, the average groundwater irrigation use is similar in both the monsoon/$Kharif$ and winter/*Rabi* season (see summary statistics for irrigation in Table 1). This implies a comparative inefficiency of irrigating in the monsoon/$Kharif$ season. Evaluation of water use efficiency has become a subject of intense debate in India. Scientific evidence drawing from satellite imagery and field studies indicate that the Indian aquifers are being depleted at an unprecedented rate relative to other countries (Famiglietti, 2014; Rodell et al., 2018; Siebert et al., 2010). The treatment of groundwater as a common resource with very little regulation, along with highly subsidised electricity for agricultural, are seen as the leading factors of mismanagement (Shah, 2013; Dubash, 2007; Badiani-Magnusson and Jessoe, 2018).

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

In response to the first channel, we analyse shifts in crop choice (reported in Columns 4 to 6 of Table 4). We consider three categories of crops – those grown in the dry winter/*Rabi* season, drought tolerant, and cash based – which are all characterised by an element of risk. The winter/*Rabi* season receives very little rainfall, hence crops grown in this season are much more vulnerable to weather shocks. Surprisingly, we find no significant shifts towards these type of crops. Drought tolerant crops are resistant to semi-arid conditions and are therefore an effective way of reducing exposure to rainfall shocks. We find that 25.4% fewer villages report one of their three most widely cultivated crops to be drought tolerant in response to an increase of one standard deviation of irrigation groundwater use. Finally, cash crops cannot be directly used for household consumption as they require post-harvest processing. While these crops are generally considered to be quite profitable they are also more susceptible to price fluctuations. For farmers who rely on their agricultural produce to feed their household, moving to cash crops can therefore often be considered risky. While our point estimate on cash crops is positive, it is not statistically significant.

We investigate the second channel by considering investments in a range of inputs (Columns 1 to 3 of Table 4). While we estimate a positive effect on ownership of mechanised farm equipment and a negative relationship with the use of water-saving technologies (such as drip and sprinklers), these are not statistically significant. However, there appears to be a strong effect of irrigation on land extensification – an increase of one standard deviation in irrigation results in an extra 15.9% of available agricultural land being cultivated.

In a further investigation, we attempt to estimate the effects of irrigation on the size of landholdings. There is good reason to believe that shocks to land productivity could have significant implications on the land market. For instance, those farmers that invest in irrigation may buy out less productive non-irrigated farms. To test for this we analyse the impact of access to irrigation on the share of households in four different landholding categories – landless, 0-2, 2-4 and above 4 acres, reported in Table 5. While we see a positive point estimate for the landless and the 0-2 acre category, this is not significantly different from zero.

B. Consumption

Table 6 reports results on the impact of irrigation on measures of consumption at the village level. Importantly, we find a 5.4% reduction in the village poverty rate for a one standard deviation increase in groundwater irrigation for agriculture (Column 2). This result agrees with other research which exploit quasi-experimental empirical approaches to analyse the effect of irrigation on poverty alleviation (Duflo and Pande, 2007; Blakeslee et al., 2020; Sekhri, 2014). This effect is further corroborated by our results on mean night light intensity, measured over a 5 year period – a one standard deviation in irrigation leads to an 18% increase in night time light (Column 3).

While our point estimate on consumption per capita is positive, it is not statistically significant (Column 1). However, we further investigate the impact of irrigation on consumption using a more direct measure – household microdata from the Socio Economic Caste Census of India (2012) on asset ownership. We find a significant positive effect on the index of household assets. A one standard deviation in groundwater irrigation increases the asset index by 0.45 standard deviations (Column 4). This effect appears to be largely driven by solid house construction – the share of households that own a solid house increases by 20% (see Table A3 which presents results on each component of the asset index independently). As explained in Duflo and Banerjee (2011), investment in solid house construction is a common form of informal savings among poor households. This implies that improved access to irrigation provides households with additional cash in hand.

C. Labour

An increase in agricultural production with improved access to 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 specialised mechanised tools such as transplanters and harvesters. Furthermore, labour demand for agriculture is likely to be influenced by market opportunities in other sectors. On-farm growth may spur production in off-farm sectors, therefore increasing demand for labour in those industries. Alternatively, irrigation may provide some villages with a comparative advantage in farming, thereby attracting labour away from other sectors or less productive population centres. Characterised by these complex interactions, the overall effect of irrigation on the labour market is ambiguous. We attempt to identify the dominance of these different components by evaluating the impact of irrigation on aggregate employment rates as well as shifts in the sectoral share of the workforce.

We begin our analysis of structural shifts in labour allocation, by analysing the effect of irrigation on the share of the population employed at the village level (reported in Panel A of Table 7, Columns 1 to 3). We find a 2.5% reduction in the share of the population employed, which is statistically significant at the 10% level (Column 1). This decline appears to be driven by a statistically significant 5.3% decline in female employment due to irrigation (Column 3). In a time series analysis of employment trends in India spanning 3 decades (from 1990 to 2010), Mehrotra et al. (2014) describe women as the reserve army of labour for the agricultural sector, called upon only in times of distress. Irrigation, which appears to significantly improve agricultural sector productivity, may therefore be reducing the need for female labour force participation. This drop may also be related to the 'income effect'; a trend suggesting that women appear to drop out of the labour force as households become wealthier (Mehrotra and Sinha, 2017).

Agriculture is the largest employer in our sample of villages, with approximately 30% of the workforce reporting their primary occupation to be either cultivation or manual agricultural labour. In order to capture sectoral shifts in labour allocation to the agricultural sector, we therefore present results on the share of the workforce employed in these categories. Results from this analysis are reported in Panel B of Table 7. We find that there is no statistically significant shifts in labour allocation either into or out of the agricultural sector. Specifically, we can only rule out a 0.44 standard deviation change in the share of the workforce employed as cultivators and a 0.40 standard deviation change in the workforce employed as manual labourers, with 95% confidence.

Data from the Population Census of India allows us to further disaggregated our outcome variables on employment in the agricultural sector by the time spent employed. Specifically, we can consider two types of workers – full-time workers are those that are economically active for more than 6 months of the year. As reported in Panel C of Table 7, we find that irrigation leads to a large and significant increase in the share of full-time workers. Specifically, we find that a one standard deviation increase in groundwater irrigation increases the share of the population in full-time employment by 10.3% (Column 1). This appears to be driven by an increase in both cultivators (Column 4) and manual labourers (Column 7). This result would indicate that there is in fact an increase in labour demand within the agricultural sector, which is absorbed by the existing labour force extending there period of activity to cover both growing seasons.

We complement this occupational data in the agricultural sector recorded by the Population Census of India, with data from the Economic Census of India on employment in village businesses, reported in Panel A of Table 8. We consider the share of the workforce employed across all village businesses, as well as in the following sectors independently: agro-processing, livestock, construction, manufacturing, and services. The point estimate for employment in the agro-processing sector (most closely aligned to agriculture) is negative, while point estimates in all the other sectors are positive. However, none are significantly different from zero at the 10% level. In the case of the services sector, we can only rule out a 0.41 standard deviation change in the dependent variable in response to a one standard deviation change in irrigation. In Panel B of Table 8, we consider the effect of irrigation on the number of business establishments in each industry. Here again we do not detect any statistically significant shifts. Taken together, it would appear that irrigation does not lead to substantive shifts in off-farm sectoral growth at the local level. These results add to evidence from two concurrent papers, showing that access to irrigation does not seem to cause significant shifts in structural transformation within villages in the context of rural India (Asher et al., 2021a; Blakeslee et al., 2021).

D. Demographics

As a final measure of interest we consider the effect of groundwater irrigation on village demographics, reported in 9. We find an increase in the village population of approximately 34.7% in response to a one standard deviation increase in irrigation (Column 1). While this is a large effect, it is imprecisely estimated and only significantly significant at the 10% level. At this test size, we can only rule out a less than 5% increase in the village population. This result is also reflected in our measure of population density. In an attempt to more directly attribute this effect to labour migration, we estimate the impact of irrigation on the share of the working age population – aged 15 to 65 years – both for the total population as well as separately for the male population. Our point estimate is negative for the working age population and positive for the male population, though neither is statistically significant.

It is possible that shifts in demographics from access to groundwater irrigation also affect nearby agglomerations. Asher et al. (2021a) find that canals cause a 20% increase in the population density of urban centres. Our measure of irrigation however, based on pump functionality with groundwater at a highly localised level, will have a much higher degree of spatial disaggregation relative to the large catchment area of a canal and therefore does not lend itself for this type of analysis.

E. Heterogeneity by Landholding Size

In this paper, we consider the effect of access to the dominant source of irrigation for Indian farmers – groundwater accounts for approximately 70% of irrigated land area across the country. This type of irrigation technology, requires significant private investment from farmers. Purchasing a centrifugal pump alone costs approximately INR 16,000 (GBP 160) which is close to the average annual per capita consumption of households in our sample of villages (see Table 1). Our decision making framework predicts that returns to investment are likely to be more meaningful for farmers with larger landholdings. Given that we reject consolidation of landholdings in response to irrigation (see Table 5), we treat ex-post observation of landholdings as a covariate upon which to conduct heterogeneity analysis.[31]

Our results from this analysis suggest that for villages with high landholding size, access to groundwater irrigation leads to consistently higher agricultural yields in the dry winter/*Rabi* season; see Columns 1 to 6 of Table 10. For each of our proxies – mean,

[31] To conduct this analysis, we split our sample evenly by villages with high or low average village landholding size, calculated as above or below the median respectively. We run regressions separately for both samples and compare our results.

maximum, and difference between early-season and the maximum – the point estimate for high landholding villages is at least twice the value of those for low landholding villages. This translates into significantly higher consumption gains among villages with larger landholdings (Column 7 and 8). However, despite incurring relatively higher returns villages with larger landholdings show no evidence of labour re-allocation between local village industries (Table 11).

6. Conclusion

First documented with regards to the Industrial Revolution in England during the 18th century, scholars argued that an increase in agricultural productivity was a necessary precursor for industrial growth (Nurkse, 1953; Robinson, 1954). Since then, a large literature has been devoted to chronicle the process of economic growth across countries, overwhelmingly finding that this is accompanied by a process of structural re-allocation of labour away from the agricultural sector towards the manufacturing and service industries (Herrendorf et al., 2014). Leveraging this information, a number of theoretical models on structural change have been formalised, placing agricultural growth as a catalyst for the process of industrialisation (Gollin et al., 2002; Ngai and Pissarides, 2007). Conversely, other models have suggested that agricultural productivity may in fact impede the process of growth if the sector has a comparative advantage; thereby pooling in labour (Matsuyama, 1992). Yet despite the centrality of this discussion in understanding the development process of low-income countries, there exists remarkably little empirical evidence to support the mechanisms suggested by these models.

In this paper, we estimate the impact of a shock to agricultural productivity – induced by access to groundwater irrigation – on the structural transformation of the rural economy in India. Primarily, we find that access to irrigation does in fact significantly boost agricultural production. Furthermore, farmers appear to significantly re-optimise their production strategies in terms of land area cultivated and types of crops grown. Secondly, this production shock has a large and positive effect on alleviation of poverty at the village level. Finally, irrigation appears to increase demand for labour in the agricultural sector – with 7% of cultivators and 10% of manual labourers shifting from part-time to full-time employment – as farmers now cultivate during the dry winter/*Rabi* season. However, it does not lead to any significant shifts in labour re-allocation between sectors.

Since the onset of the Green Revolution in India during the 1960s, the Government has adopted policies of providing free or largely subsidised electricity to farmers for irrigation in an effort to stimulate growth in the rural economy. Subsequently, the area irrigated by groundwater has increased by over 500 percent between 1960 and 2010, now accounting for close to 70% of the cultivated land area (Foster and Garduño, 2013). Exploitation of this resource however, is now becoming a growing concern – India has the fastest depleting aquifers in the world (Famiglietti, 2014). Given the seriousness and extent of these negative externalities from promoting groundwater use, they should be balanced with reliable evidence on the benefits of irrigation as an instrument to stimulate rural economic growth. The results from this paper suggest the irrigation in the monsoon/$Kharif$ season is relatively inefficient in terms of yield benefits and should therefore be regulated. The major benefits are accrued during the dry winter/*Rabi* season by allowing a reliable second cropping season. This increase in production is coupled with significant asset accumulation, reductions in poverty, and stabilisation of inter-annual economic activity within the village.

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," 2021.

_ 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," 2021. World Bank Economic Review.

Badiani-Magnusson, Reena and Katrina Jessoe, "Electricity Prices, Groundwater, and Agriculture: The Environmental and Agricultural Impacts of Electricity Subsidies in India," in "Agricultural Productivity and Producer Behavior," University of Chicago Press, 2018, pp. 157–183.

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 Pelegrina, and Singh Karan, "Irrigation and the Spatial Pattern of Structural Transformation in India," 2021.

_ , 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.

Carillo, Mario F, "Agricultural policy and long-run development: evidence from Mussolini's Battle for Grain," *The Economic Journal*, 2021, *131* (634), 566–597.

Córdoba, Juan Carlos and Marla Ripoll, "Agriculture and aggregation," *Economics Letters*, 2009, *105* (1), 110–112.

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 Abhijit Banerjee, *Poor economics*, PublicAffairs, 2011.

_ **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.

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.

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.

Field, Alexander James, "Sectoral shift in antebellum Massachusetts: A reconsideration," *Explorations in Economic History*, 1978, *15* (2), 146–171.

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

Foster, Stephen and Héctor Garduño, "Groundwater-resource governance: are governments and stakeholders responding to the challenge?," *Hydrogeology Journal*, 2013, *21* (2), 317–320.

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 L Parente, and Richard Rogerson, "The food problem and the evolution of international income levels," *Journal of Monetary Economics*, 2007, *54* (4), 1230–1255.

_ , 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.

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.

Johnston, Bruce F, "Agriculture and structural transformation in developing countries: A survey of research," *Journal of Economic Literature*, 1970, *8* (2), 369–404.

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie, "Beyond balanced growth," *The Review of Economic Studies*, 2001, *68* (4), 869–882.

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.

Kuznets, Simon and John Thomas Murphy, *Modern economic growth: Rate, structure, and spread*, Vol. 2, Yale University Press New Haven, 1966.

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.

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.

Mkhabela, Manasah S, Milton S Mkhabela, and Nkosazana N Mashinini, "Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR," *Agricultural and Forest Meteorology*, 2005, *129* (1-2), 1–9.

Mukherji, Aditi, Stuti Rawat, and Tushaar Shah, "Major insights from India's minor irrigation censuses: 1986-87 to 2006-07," *Economic and Political Weekly*, 2013, pp. 115–124.

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.

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.

Rodell, Matthew, JS Famiglietti, DN Wiese, JT Reager, HK Beaudoing, Felix W Landerer, and M-H Lo, "Emerging trends in global freshwater availability," *Nature*, 2018, *557* (7707), 651–659.

Ryan, Nicholas and Anant Sudarshan, "Rationing the commons," Technical Report, National Bureau of Economic Research 2020.

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, Mihir, "Water: Towards a paradigm shift in the twelfth plan," *Economic and Political weekly*, 2013, pp. 40–52.

Siebert, Stefan, Jacob Burke, Jean-Marc Faures, Karen Frenken, Jippe Hoogeveen, and Felix Theodore Portmann, "Groundwater use for irrigation—a global inventory," *Hydrology and Earth System Sciences*, 2010, *14* (10), 1863–1880.

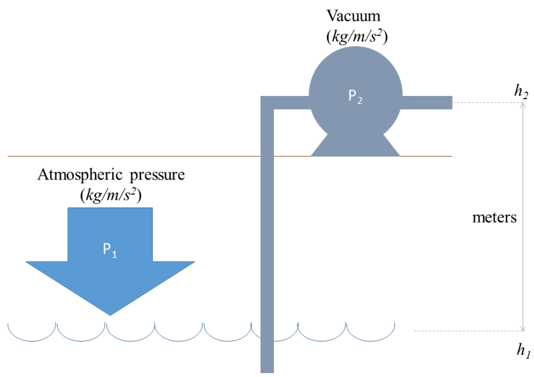
Simonsen, Marianne, Lars Skipper, and Niels Skipper, "Price sensitivity of demand for prescription drugs: exploiting a regression kink design," *Journal of Applied Econometrics*, 2016, *31* (2), 320–337.

Syrquin, Moshe, "Patterns of structural change," *Handbook of development economics*, 1988, *1*, 203–273.

Vollrath, Dietrich, "The agricultural basis of comparative development," *Journal of Economic Growth*, 2011, *16* (4), 343–370.

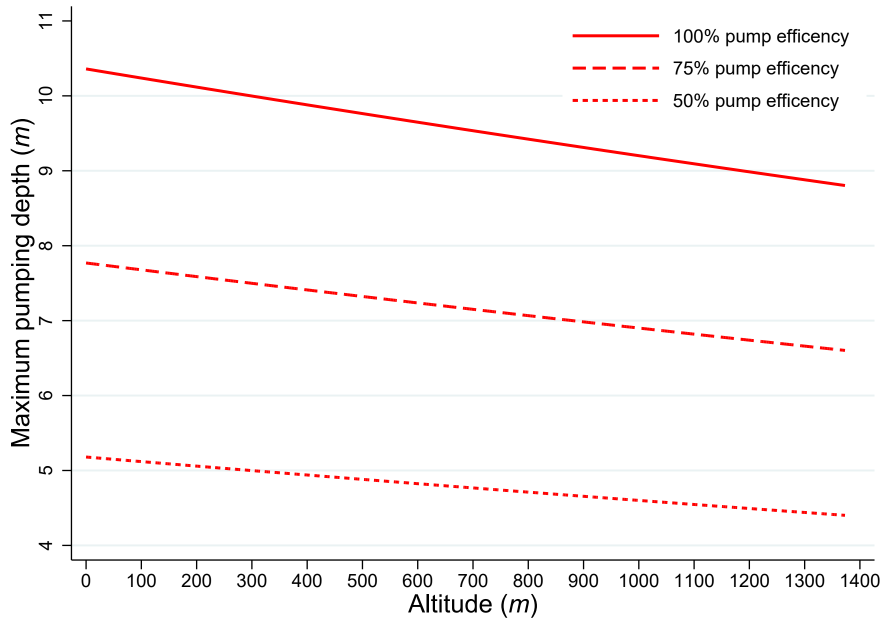
Wardlow, Brian D and Stephen L Egbert, "A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas," *International Journal of Remote Sensing*, 2010, *31* (3), 805–830.

Figure 1: Pumping mechanism of a centrifugal pump



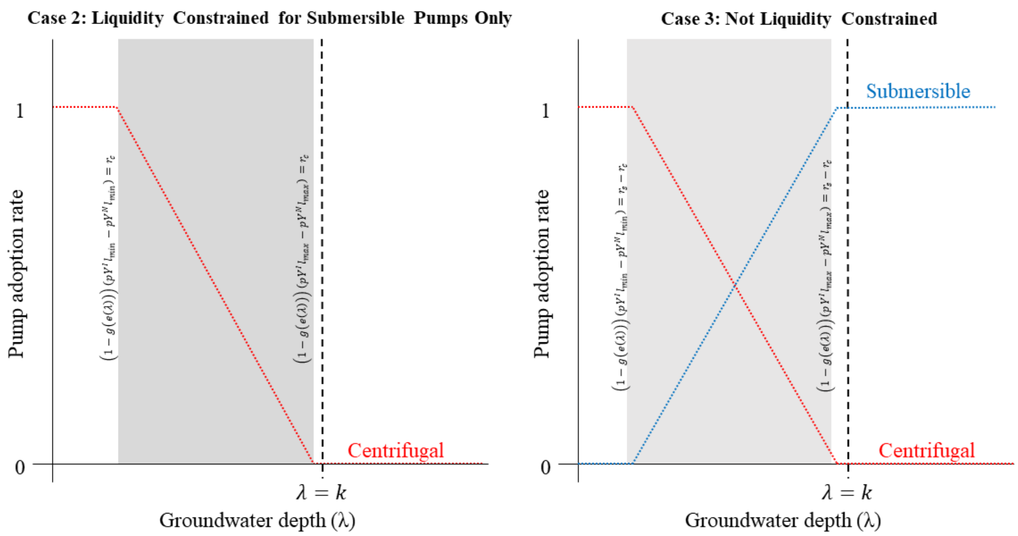
Notes: Extraction of water from a well using a centrifugal pump can be described by Bernoulli's principle of fluid dynamics (Equation 1). Assuming constant flow velocity, water extraction is defined by Equation 2: $h_2 - h_1 = \frac{P_1 - P_2}{\rho g}$, where P_1 and P_2 refers to pressure from the water table and the pump respectively, ρ is the density of water (997 kg/m^3), and g the gravitational force (9.81 m/s^2). The difference between h_1 and h_2 is the distance between the ground level and the water table. Under a perfect vacuum (P_2=0 $kg/m/s^2$) and atmospheric pressure at sea-level (P_1=101,325 $kg/m/s^2$), the depth from which water can be extracted is 10.33 meters.

Figure 2: Effect of efficiency and altitude on the maximum pumping depth of a centrifugal pump



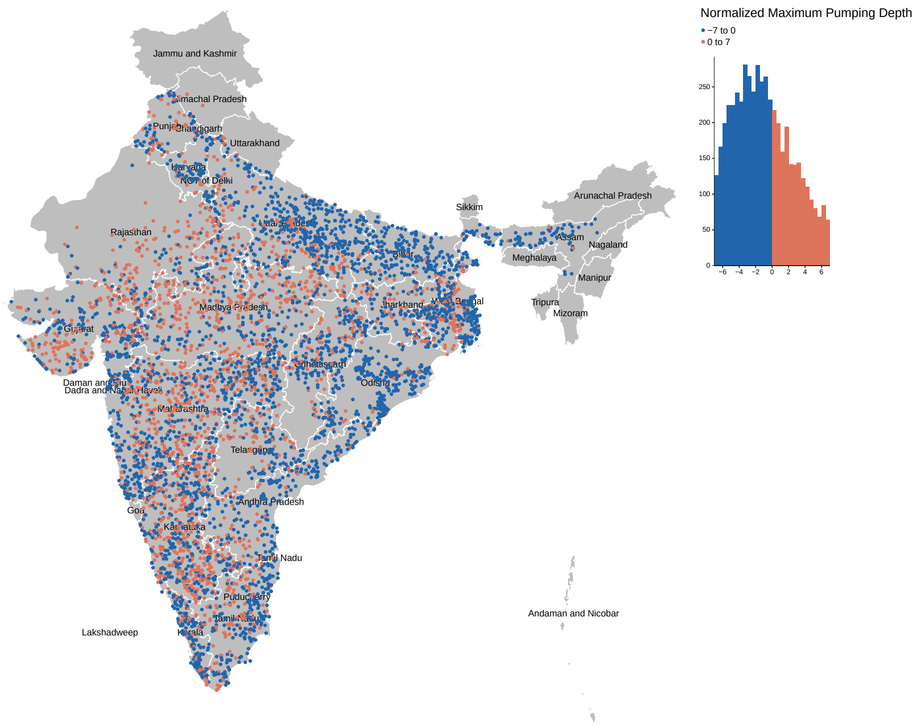
Notes: 100% pump efficiency occurs under a perfect vacuum (where P=0 $kg/m/s^2$). 75% and 50% pump efficiency indicates the corresponding percentage drop from a case of perfect vacuum. The range of altitude plotted correspond to those found in our sample of villages. Data on altitude was extracted from raster files for the whole of India, obtained from the ALOS Global Digital Surface Model. A barometric formula was used to calculate atmospheric pressure with 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.

Figure 3: Illustrative diagram for the evolution of pump adoption with groundwater depth



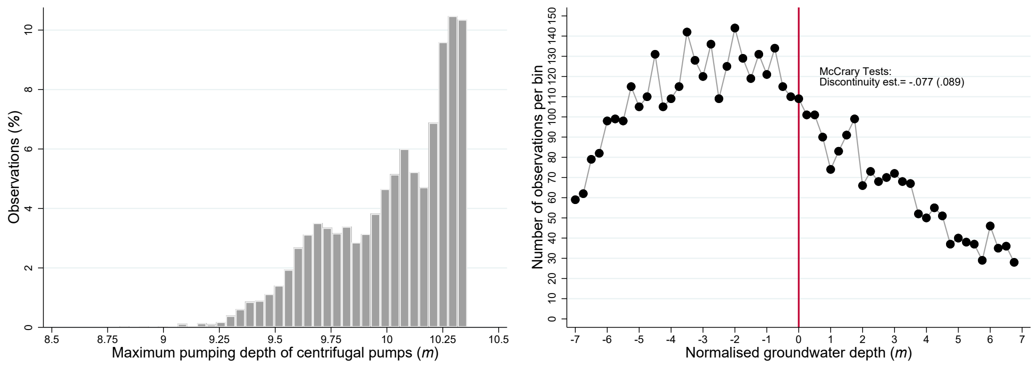
Notes: It is the subset of farmers that can afford a centrifugal pump but not a submersible (i.e. Case 2) that generates a kink in overall pump adoption and consequently irrigation. The linear functional form requires that pump efficiency and land are distributed uniformly. Additionally, under the assumption that $g(e(\lambda_v)) = 0$ – that is, centrifugal pumps work with certainty – even the smallest farmers (whose decision equation is indexed by min) will adopt centrifugal pumps. On the other hand, as $\lambda_v - k$ becomes very small, only the farmers with the largest landholdings (whose decision equation is indexed by max) are most likely to adopt centrifugal pumps.

Figure 4: Location of sample villages and their groundwater depth



Notes: Our total sample covers 4,896 villages across 477 districts amongst 19 States. Each point on the map represents a village in this sample. The maximum groundwater depth below which no centrifugal pump can operate is 10.33 meters. Red points correspond to villages with depths deeper than 10.33 metres, while blue points are shallower.

Figure 5: Distribution of the assignment variable across the kink point

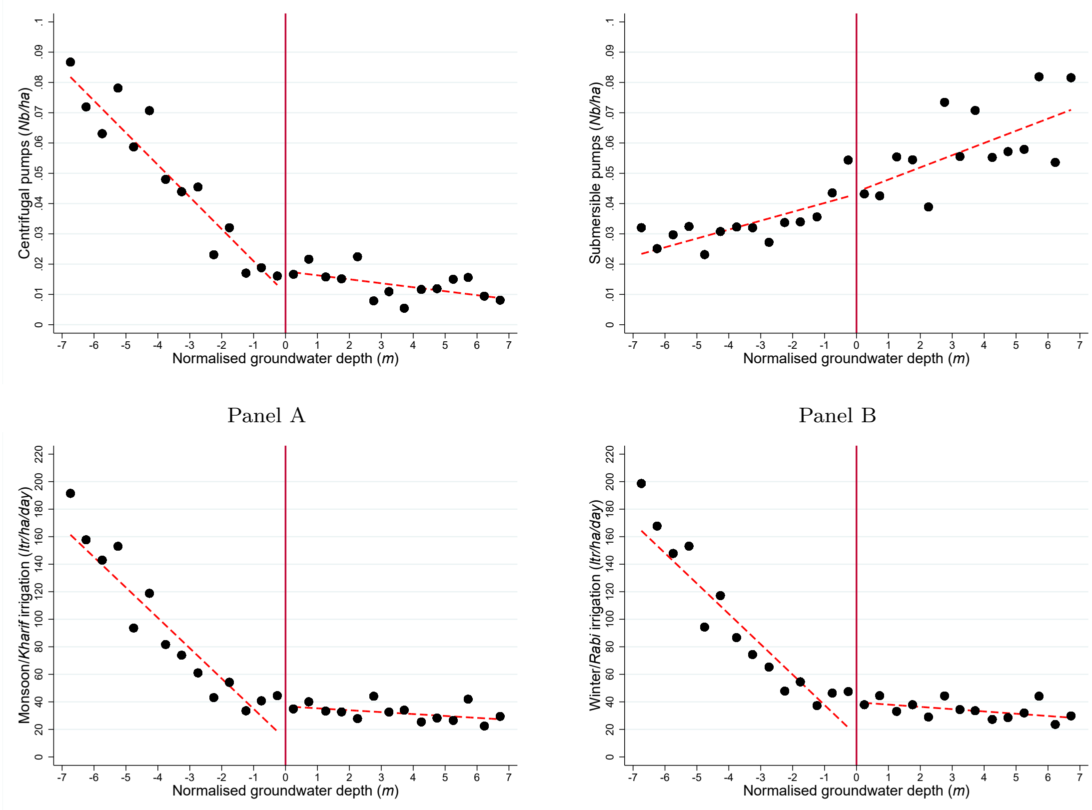


Panel A

Panel B

Notes: The kink point refers to the groundwater depth below which no centrifugal pump can operate – 10.33 meters. Panel A shows the distribution of the maximum pumping depth of a centrifugal pump for villages in our sample. Panel B plots the number of observations in each bin for groundwater depth normalised at the kink point. 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 test by estimating the log change in height between bins at the kink point.

Figure 6: Deterministic relation between groundwater depth and irrigation



Panel A

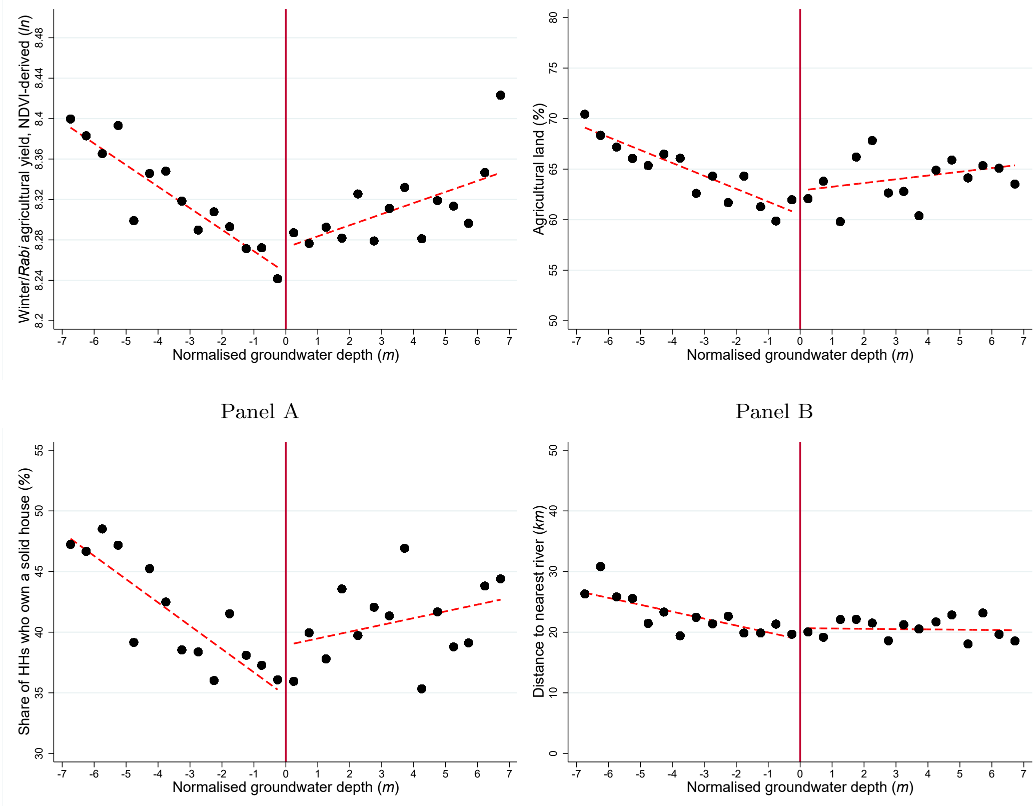
Panel B

Panel C

Panel D

Notes: The x-axis in each panel represents our assignment variable – the maximum groundwater depth recorded at any point between 2011-2013. This variable is normalised around the kink point of 10.33 metres – the operational threshold for a centrifugal pump. Points to the right of zero correspond to depths deeper than 10.33 metres, 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. Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal and submersible pumps in Panels A and B respectively. Irrigation, calculated as water input in litres (for a complete discussion on the construction of this variable see Appendix B), is reported for the monsoon/*Kharif* (June-October) and the dry winter/*Rabi* season (November-March) in Panels C and D respectively. 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. Formal estimates of the kink for these variables using *fuzzy* RK regression analysis are reported in Table 2.

Figure 7: Deterministic relation between groundwater depth and a selection of outcomes and covariates



Panel A

Panel B

Panel C

Panel D

Notes: The x-axis in each panel represents our assignment variable – the maximum groundwater depth recorded at any point between 2011-2013. This variable is normalised around the kink point of 10.33 metres – the operational threshold for a centrifugal pump. Points to the right of zero correspond to depths deeper than 10.33 metres, 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 and covariates. We present graphical evidence on three outcome variables: winter/*Rabi* agricultural yield (derived from NDVI – an index of vetegation cover based on satellite imagery, for a complete discussion on the construction of this variable see Appendix B) calculated as the log of the difference between early-season and the maximum value (Panel A), agricultural land calculated as the percentage share of village area used for agricultural purposes (Panel B), and the percentage share of households in the village that own a solid house (Panel C). Additionally we present graphical evidence for one covariate variable: distance, measured in kilometres, to the closest river (Panel D). 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. Formal estimates of the kink for these variables using *fuzzy* RK regression analysis are reported in Tables 3, 4, A3, and 2 for Panels A, B, C and D respectively.

Table 1: Descriptive statistics

	Mean (1)	SD (2)	N (3)	Source (Year) (4)
Panel A: Irrigation				
Monsoon/*Kharif* irrigation (*ltr/ha/day*)	65.494	141.903	4896	MIC (2013)
Winter/*Rabi* irrigation (*ltr/ha/day*)	68.121	144.894	4896	MIC (2013)
Irrigation average (*ltr/ha/day*)	66.873	142.736	4896	MIC (2013)
Tube-wells (*nb/ha*)	0.083	0.147	4896	MIC (2013)
Centrifugal pumps (*nb/ha*)	0.034	0.092	4896	MIC (2013)
Share of agricultural area irrigated by tube-wells (%)	70.122	45.650	4896	MIC (2013)
Panel B: Demographics and amenities				
Population (*nb*)	4253.026	5261.223	4896	PC (2011)
Share of population is literate (%)	61.349	11.709	4896	PC (2011)
Share of population from scheduled castes (%)	18.160	15.677	4896	PC (2011)
Panel C: Consumption				
Per capita consumption ('000 *Rs./annum*)	18.070	4.717	4896	SECC (2012)
Share of households are BPL[a] (%)	30.930	18.362	4896	SECC (2012)
Share of households who own a solid house (%)	40.716	28.570	3600	SECC (2012)
Panel D: Agriculture				
Average village landholding size (*ha*)	3.427	5.746	3600	SECC (2012)
Share of households who own mechanised equipment (%)	4.078	6.894	3600	SECC (2012)
Share of workforce are cultivators (%)	13.319	10.239	4896	PC (2011)
Share of workforce are labourers (%)	18.394	11.951	4896	PC (2011)
Panel E: Industry				
Number of establishments (*nb*)	249.779	457.944	4896	EC (2012)
Share of workforce employed (%)	21.247	19.165	4896	EC (2012)
Share of workforce in agro-processing (%)	0.262	1.090	4896	EC (2012)
Share of workforce in manufacturing (%)	3.606	5.904	4896	EC (2012)
Share of employment in services (%)	8.824	9.262	4896	EC (2012)
Share of employment in construction (%)	0.263	0.724	4896	EC (2012)

Notes: For additional details on the source of data and construction of each variable, refer to Section 3. The total sample with non-missing observations across all our outcomes of interest and within our bandwidth (7 metres) covers 4,896 villages across 477 districts in 19 States. Variables obtained from the SECC however, have a slightly reduced sample covering 3,600 villages. [a]Poverty line is set at Rs.31/day.

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

	Pump adoption		Groundwater irrigation			Covariates		
	Centrifugal pumps (*nb/ha*)	Submersible pumps (*nb/ha*)	Average (*standardised*)	Monsoon *Kharif* (*standardised*)	Winter *Rabi* (*standardised*)	Distance to nearest river (*km*)	Inside a canal command area (*binary*)	Potential yield (*index*)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
π_1	0.003***	0.000	0.090***	0.091***	0.087***	0.332	0.006	0.013
	(0.001)	(0.001)	(0.012)	(0.012)	(0.012)	(0.329)	(0.004)	(0.012)
Mean	0.034	0.041	-0.000	-0.000	0.000	21.783	0.081	-0.081
SD	0.092	0.095	1.000	1.000	1.000	23.185	0.273	0.963
N	4896	4896	4896	4896	4896	4896	4896	4896

Notes: This table presents estimates on the effect of groundwater depth on pump adoption, irrigation, and covariates. π_1 is the estimated change in slope of the assignment rule at the kink (based on Equation 9). Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal (Column 1) and submersible (Column 2) pumps. We calculate irrigation as water input in litres. However, for the purpose of interpretation across all regressions, we standardise these variables (for a complete discussion on the construction of these variables, see Appendix B). Our measure of irrigation is reported in Columns 3 to 5 as an average over the year, as well as independently for the Monsoon/*Kharif* (June-October) and the dry Winter/*Rabi* season (November-March) respectively. We consider three covariates (reported in Columns 6 to 8) capturing village-level ecological endowment variables unrelated to irrigation. Distance to the nearest river captures the minimum distance, measured in kilometres, to the closest river. Inside a canal command area is a binary variable for whether the village has tube-wells located in the command area of a dam/canal irrigation network. Maximum potential yield is an index calculated using the agro-ecological zones' potential yield for 15 crops under rain-fed conditions. All regressions include state dummies. Heteroskedasticity robust standard errors are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3: Impact of groundwater irrigation on agricultural yield

	Monsoon/*Kharif*			Winter/*Rabi*		
	Differenced (*ln*)	Mean (*ln*)	Max (*ln*)	Differenced (*ln*)	Mean (*ln*)	Max (*ln*)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Agricultural yield, NDVI-derived						
Irrigation	0.036	0.050**	0.070***	0.098***	0.012	0.084***
(*standardised*)	(0.062)	(0.023)	(0.027)	(0.035)	(0.021)	(0.030)
Mean	3411.471	4331.142	6621.835	253.350	5022.151	5273.851
SD	1229.687	837.629	1065.302	1575.954	1200.486	1266.292
Panel B: Agricultural yield, EVI-derived						
Irrigation	0.037	0.033	0.042	-0.009	0.008	0.005
(*standardised*)	(0.068)	(0.026)	(0.034)	(0.038)	(0.018)	(0.028)
Mean	2400.768	3049.306	4603.180	1734.291	2938.172	4671.421
SD	1014.689	628.628	966.622	959.040	759.095	1055.848
N	4896	4896	4896	4896	4896	4896

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on agricultural yield. Irrigation intensity is calculated as *litres/ha/day* and standardised. We rely on measures of vegetation cover using satellite imagery – NDVI (reported in Panel A) and EVI (reported in Panel B) – as indicators for agricultural yield. Calculated over a three year period (2011-2013), we consider three proxies specific to each agricultural season – maximum, mean, and the difference between early season and the maximum value. The Monsoon/*Kharif* season (reported in Columns 1 to 3) is based on data from June to October, and the dry Winter/*Rabi* season (reported in Columns 4 to 6) covers November to March. For a complete discussion on the data and how each proxy is calculated refer to Appendix B. Summary statistics for all our proxies are reported on the level form of the variables. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 4: Impact of groundwater irrigation on agricultural production choices

	Inputs			Crop choice		
	Mechanised equipment (%) (1)	Water-saving technology (%) (2)	Agricultural land (%) (3)	Winter/ *Rabi* (*binary*) (4)	Drought tolerant (*binary*) (5)	Cash (*binary*) (6)
Irrigation	1.208	-4.493	15.972***	0.071	-0.254***	0.050
(*standardised*)	(1.558)	(2.794)	(4.580)	(0.062)	(0.084)	(0.072)
Mean	4.078	3.228	64.364	0.242	0.315	0.198
SD	6.894	15.485	25.330	0.429	0.465	0.398
N	3600	4896	4896	3848	3848	3848

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on agricultural production choices. Irrigation intensity is measured as *litres/ha/day* and standardised. Columns 1 to 3 considers the effect of irrigation on three measures of inputs: mechanisation – calculated as the percentage share of households who own mechanised farm equipment (e.g. tractors, harvesters etc.), water-saving technology – calculated as the percentage share of tube-wells which are adapted to water-saving mechanisms (i.e. which use drip and sprinklers), and agricultural land – calculated as the percentage share of village area used for agricultural purposes. Columns 4 to 6 present estimates on the effect of irrigation on three binary measures of crop choice: does a village grow winter crops (wheat, barley, potato, oilseed, and chickpea), drought tolerant crops (millet, sorghum, maize, pigeon pea, and groundnut), or cash crops (sugarcane, oilseed, cotton, and tobacco). All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 5: Impact of groundwater irrigation on the distribution of landholdings

	Landless (%) (1)	0-2 Acres (%) (2)	2-4 Acres (%) (3)	4+ Acres (%) (4)
Irrigation	1.522	0.491	-2.543	0.601
(*standardised*)	(4.699)	(3.746)	(1.575)	(2.587)
Mean	55.696	22.641	9.271	12.081
SD	22.904	19.078	7.296	12.756
N	3600	3600	3600	3600

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on the distribution of landholdings. Irrigation intensity is measured as *litres/ha/day* and standardised. Results are reported for four categories of land acreage – 0, 0-2, 2-4, and over 4. Each variable is calculated as the percentage share of households who own that specific landholding size. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 6: Impact of groundwater irrigation on consumption

	Consumption per capita (*ln*) (1)	Poverty rate (*share*) (2)	Mean night light (*ln*) (3)	Household assets (*index*) (4)
Irrigation	0.060	-0.054*	0.178**	0.451*
(*standardised*)	(0.039)	(0.028)	(0.081)	(0.230)
Mean	18.215	0.316	6.715	0.317
SD	4.900	0.190	4.665	0.994
N	4896	4896	4896	3600

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on consumption. Irrigation intensity is measured as *litres/ha/day* and standardised. We consider two household level measures of consumption (reported in Columns 1 and 2): imputed log consumption per capita and the share of the population living below the poverty line (set at Rs.31/day). For a complete discussion on the data and construction of these variables refer to Appendix B. Additionally, we rely on measures of night light luminosity from satellite images as a proxy for consumption (reported in Column 3). We calculate the average of mean night light over a five year period (2009-2013). For a complete discussion on the data and construction of this variables refer to Appendix B. Finally, household asset ownership (reported in Column 4) is an 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 (2012). Table A3 in Appendix A presents results for the effect of irrigation on each asset independently. Summary statistics for consumption per capita, as well as those for night light, are reported on the level form of the variable. All regressions include state dummies. 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 7: Impact of groundwater irrigation on agricultural sector employment

	Total			Cultivators			Manual labourers		
	Person (%) (1)	Male (%) (2)	Female (%) (3)	Person (%) (4)	Male (%) (5)	Female (%) (6)	Person (%) (7)	Male (%) (8)	Female (%) (9)
Panel A: Share of working age population									
Irrigation	-2.481*	0.404	-5.316**	0.506	1.599	-0.614	-1.701	-1.035	-2.576
(*standardised*)	(1.464)	(0.903)	(2.550)	(1.605)	(1.862)	(1.674)	(1.850)	(1.839)	(2.215)
Mean	44.582	55.390	33.280	13.319	18.250	8.176	18.394	18.915	17.786
SD	10.354	6.539	17.372	10.239	11.718	10.727	11.951	11.637	14.588
Panel B: Share of workforce									
Irrigation	-	-	-	2.521	2.957	0.557	-1.849	-1.905	0.285
(*standardised*)	(-)	(-)	(-)	(3.099)	(3.244)	(3.408)	(3.393)	(3.166)	(4.319)
Mean	-	-	-	29.224	32.873	21.774	39.853	33.957	49.613
SD	-	-	-	19.585	20.428	21.570	21.502	20.032	27.084
Panel C: Share of full-time workers									
Irrigation	10.289***	8.637***	8.019*	6.648**	5.347*	2.307	10.224**	8.046	9.763*
(*standardised*)	(3.583)	(3.147)	(4.610)	(3.239)	(2.916)	(5.074)	(4.971)	(4.902)	(5.421)
Mean	73.651	81.052	59.401	84.616	89.091	67.317	61.758	68.384	52.010
SD	21.592	18.970	29.319	20.191	18.401	32.115	30.915	30.358	34.540
N	4896	4896	4896	4896	4896	4896	4896	4896	4896

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on employment of the village population in the agricultural sector. Irrigation intensity is measured as $litres/ha/day$ and standardised. Alongside total employment (reported in Columns 1 to 3), we consider two specific occupational categories in agriculture: cultivators (reported in Columns 4 to 6) are those who cultivate their own land, and manual labourers (reported in Columns 7 to 9) are those who work for a daily wage. Furthermore, we disaggregate each of our categories by gender. Panel A reports results on the percentage share of the working age population employed, calculated - for total employment as well as specific to each category - as the ratio of those employed to the total population of working age. Panel B reports results on the percentage share of the workforce engaged in each category, calculated as the ratio of those employed in that category 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 - for total employment as well as specific to each category - as the ratio of full-time workers to the total workforce. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 8: Impact of groundwater irrigation on industrial sectoral employment and number of establishments

	Total (1)	Agro-processing (2)	Livestock (3)	Construction (4)	Manufacturing (5)	Services (6)
Panel A: Share of workforce (%)						
Irrigation	0.983	-0.235	1.167	0.150	0.026	0.313
(*standardised*)	(3.043)	(0.180)	(1.617)	(0.120)	(1.009)	(1.479)
Mean	21.247	0.262	4.924	0.263	3.606	8.824
SD	19.165	1.090	9.572	0.724	5.904	9.262
Panel B: Establishments (*nb*)						
Irrigation	83.095	-0.644	19.087	1.565	2.570	41.841
(*standardised*)	53.002	1.247	22.069	1.177	10.703	28.921
Mean	249.779	2.076	68.078	2.612	38.548	122.033
SD	457.944	7.389	171.334	7.444	79.370	226.965
N	4896	4896	4896	4896	4896	4896

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on employment and number of business establishments within the village. Irrigation intensity is measured as *litres/ha/day* and standardised. Panel A reports results on sectoral employment. This is measured as the percentage share of the workforce employed, both in total for all village businesses (Column 1) as well as in each of the following sectors independently: agro-processing (this excludes crop production), livestock, construction, manufacturing, and services (Columns 2 to 6 respectively). Panel B reports results on the number of establishments. This is also reported in total for all village businesses, as well as for each industry independently. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 9: Impact of groundwater irrigation on the village population

	Population	Population density	Working age population	Male working age population
	(*ln*)	(*ln*)	(%)	(%)
	(1)	(2)	(3)	(4)
Irrigation	0.347*	0.445***	-0.455	0.676
(*standardised*)	(0.179)	(0.147)	(0.838)	(1.188)
Mean	7.783	5.888	61.295	71.144
SD	1.105	0.951	4.744	6.849
N	4896	4896	3600	3600

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on village population. Irrigation intensity is measured as *litres/ha/day* and standardised. We consider four measures of population - log of population in 2011, log of population density in 2011, share of the population that is of working age (15 to 65 years), share of the male population that is of working age (15 to 65 years). All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 10: Impact of groundwater irrigation on measures of agricultural yield and consumption by landholding size

	Winter/*Rabi* agricultural yield, NDVI-derived						Consumption	
	Differenced (*ln*)		Mean (*ln*)		Max (*ln*)		Mean night light (*ln*)	
	High (1)	Low (2)	High (3)	Low (4)	High (5)	Low (6)	High (7)	Low (8)
Irrigation	0.159	0.079	0.082	0.015	0.233**	0.069	0.496**	0.072
(*standardised*)	(0.108)	(0.055)	(0.069)	(0.032)	(0.108)	(0.044)	(0.234)	(0.135)
Mean	86.272	191.609	4900.922	5162.482	4985.818	5351.233	6.321	6.609
SD	1485.692	1443.166	1138.975	1196.587	1266.923	1211.838	4.090	4.667
N	1772	1773	1772	1773	1772	1773	1772	1773

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on agricultural yield and consumption by landholding size. The sample is split by villages having high or low landholding size, calculated as the village average landholding being above or below the median respectively. Irrigation intensity is calculated as *litres/ha/day* and standardised. We rely on measures of vegetation cover using satellite imagery – NDVI – as indicators for agricultural yield during the dry winter/*Rabi* season (November to March). Calculated over a three year period (2011-2013), we consider three proxies – maximum, mean, and the difference between early season and the maximum value (reported in Columns 1 to 6). For a complete discussion on the data and how each proxy is calculated refer to Appendix B. Additionally, we rely on measures of night light luminosity from satellite images as a proxy for consumption (reported in Columns 7 and 8). We calculate average of mean night light over a five year period (2009-2013). For a complete discussion on the data and construction of these variables refer to Appendix B. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Table 11: Impact of groundwater irrigation on the sectoral share of the workforce by landholding size

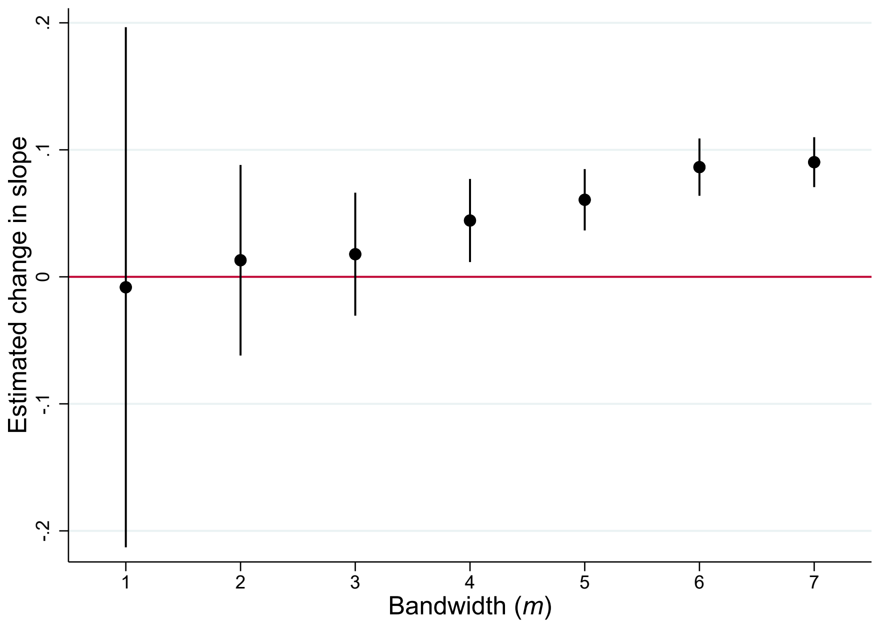
	Cultivators (%)		Manual labourers (%)		Construction (%)		Manufacturing (%)		Services (%)	
	High (1)	Low (2)	High (3)	Low (4)	High (5)	Low (6)	High (7)	Low (8)	High (9)	Low (10)
Irrigation	-2.794	5.041	1.908	0.378	-0.933	0.366	0.135	0.026	-2.340	2.375
(*standardised*)	(9.171)	(4.882)	(10.066)	(5.230)	(2.670)	(1.739)	(0.322)	(0.187)	(4.056)	(2.600)
Mean	30.908	28.035	43.438	36.526	3.086	4.217	0.220	0.295	7.714	9.951
SD	20.131	19.305	21.478	20.898	5.440	6.231	0.626	0.774	7.939	10.006
N	1772	1773	1772	1773	1772	1773	1772	1773	1772	1773

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on the sectoral share of the workforce by landholding size. The sample is split by villages having high or low landholding size, calculated as the village average landholding being above or below the median respectively. Irrigation intensity is measured as $litres/ha/day$ and standardised. We report results for 5 sectors: cultivators, manual labourers, manufacturing, construction, and services. The percentage share of the workforce engaged in each sector is calculated as the ratio of those employed in that sector to the total workforce. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

Appendices

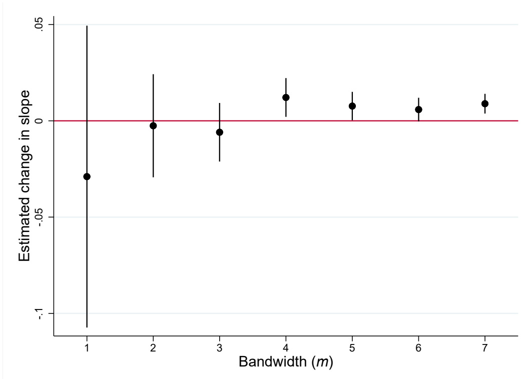
A. Additional Tables and Figures

Figure A1: Estimated kink in the deterministic relation between irrigation and groundwater depth at a range of bandwidths

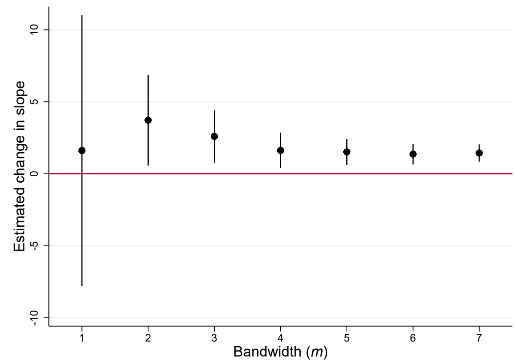


Notes: The plot presents point estimates and 90% confidence intervals for the linear specification of Equation 9 on our measure of irrigation at one meter interval bandwidths. Irrigation, is calculated as water input in litres measured as an average over the year and standardised for the purpose of all regressions (for a complete discussion on the construction of this variable see Appendix B). The regression is estimated using heteroskedasticity robust standard errors and includes state dummies.

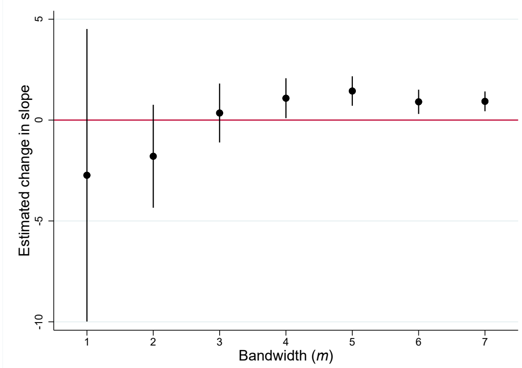
Figure A2: Estimated kink in the relation between key outcome variables and groundwater depth at a range of bandwidths



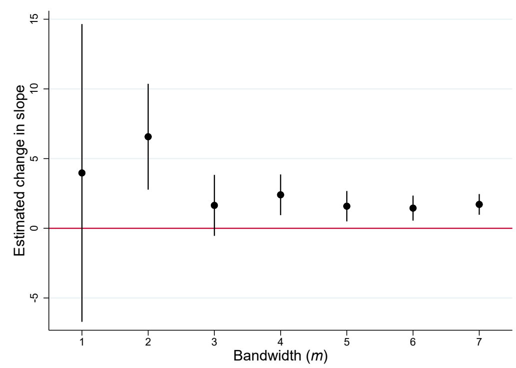
Panel A: Winter/*Rabi* agricultural yield, NDVI-derived (*ln*)



Panel B: Agricultural land (%)



Panel C: Share of full-time workers (persons) (%)



Panel D: Share of HHs who own a solid house (%)

Notes: Each plot presents point estimates and 90% confidence intervals for the linear specification of Equation 8 at one meter interval bandwidths for a selection of our outcome variables. The four outcome variables reported here include: winter/*Rabi* agricultural yield (derived from NDVI – an index of vetegation cover based on satellite imagery) calculated as the log of the difference between early-season and the maximum value (Panel A), agricultural land calculated as the percentage share of village area used for agricultural purposes (Panel B), percentage share of full-time workers (those that work for more than 6 months of the year) to the total workforce (Panel C), and the percentage share of households in the village that own a solid house (Panel D). All regressions are estimated using heteroskedasticity robust standard errors and include state dummies.

Table A1: Estimated kink in the deterministic relation of groundwater depth with pump adoption, groundwater irrigation, and covariates for different functional forms

	Pump adoption		Groundwater irrigation			Covariates		
	Centrifugal pumps (*nb/ha*)	Submersible pumps (*nb/ha*)	Average (*standardised*)	Monsoon *Kharif* (*standardised*)	Winter *Rabi* (*standardised*)	Distance to nearest river (*km*)	Inside a canal command area (*binary*)	Potential yield (*index*)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Linear								
π_1	0.003***	0.000	0.090***	0.091***	0.087***	0.332	0.006	0.013
	(0.001)	(0.001)	(0.012)	(0.012)	(0.012)	(0.329)	(0.004)	(0.012)
Panel B: Quadratic								
π_1	0.008**	-0.003	-0.036	-0.020	-0.049	-0.053	-0.010	-0.035
	(0.004)	(0.005)	(0.042)	(0.042)	(0.042)	(1.164)	(0.015)	(0.042)
Panel C: Cubic								
π_1	-0.003	-0.006	-0.162*	-0.161*	-0.164*	-0.656	-0.007	0.071
	(0.008)	(0.013)	(0.095)	(0.094)	(0.096)	(2.802)	(0.036)	(0.100)
N	4896	4896	4896	4896	4896	4896	4896	4896

Notes: TThis table presents estimates on the effect of groundwater depth on pump adoption, irrigation, and covariates for varying functional form. π_1 is the estimated change in slope of the assignment rule at the kink (based on Equation 9). Panel A presents estimates using a linear functional form, Panel B is quadratic, and Panel C is cubic. We calculate irrigation as water input in litres. However, for the purpose of interpretation across all regressions, we standardise these variables (for a complete discussion on the construction of these variables, see Appendix B). Pump adoption, calculated as the number of pumps per agricultural land area, is reported for centrifugal (Column 1) and submersible (Column 2) pumps. Our measure of irrigation is reported in Columns 3 to 5 as an average over the year, as well as independently for the Monsoon/*Kharif* (June-October) and the dry Winter/*Rabi* season (November-March) respectively. We consider three covariates (reported in Columns 6 to 8) capturing village-level ecological endowment variables unrelated to irrigation. Distance to the nearest river captures the minimum distance, measured in kilometres, to the closest river. Inside a canal command area is a binary variable for whether the village has tube-wells located in the command area of a dam/canal irrigation network. Maximum potential yield is an index calculated using the agro-ecological zones' potential yield for 15 crops under rain-fed conditions. All regressions include state dummies. 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 irrigation over varying time horizons

	Irrigation		
	Average (*standardised*) (1)	Monsoon/*Kharif* (*standardised*) (2)	Winter/*Rabi* (*standardised*) (3)
Panel A: 1 year			
π_1	0.026***	0.029***	0.022**
	(0.009)	(0.009)	(0.009)
Panel B: 3 years			
π_1	0.090***	0.091***	0.087***
	(0.012)	(0.012)	(0.012)
Panel C: 5 years			
π_1	0.088***	0.088***	0.086***
	(0.013)	(0.013)	(0.013)
N	4896	4896	4896

Notes: This table presents estimates on the effect of groundwater depth on irrigation for varying time horizons. π_1 is the estimated change in slope of the assignment rule at the kink (based on Equation 9). The assignment variable is defined as the maximum groundwater depth recorded at any point over a 1, 3, and 5 year time horizon preceding 2013. We calculate irrigation as water input in litres. However, for the purpose of interpretation across all regressions, we standardise these variables (for a complete discussion on the construction of these variables, see Appendix B). Our measure of irrigation reported in Columns 1 to 3 is an average over the year, as well as independently for the Monsoon/*Kharif* (June-October) and the dry Winter/*Rabi* season (November-March) respectively. All regressions include state dummies. Heteroskedasticity robust standard errors are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table A3: Impact of groundwater irrigation on ownership of assets

	Solid house (%) (1)	Refrigerator (%) (2)	Vehicle (%) (3)	Phone (%) (4)
Irrigation	19.991***	2.150	3.524	5.808
(*standardised*)	(6.217)	(2.377)	(3.125)	(4.356)
Mean	40.716	8.799	19.707	68.053
SD	28.570	12.957	15.411	24.972
N	3600	3600	3600	3600

Notes: This table presents *fuzzy* RK estimates on the effect of groundwater irrigation on ownership of assets. Irrigation intensity is measured as *litres/ha/day* and standardised. Columns 1 to 4 present estimates on the percentage share of households in the village that own each of the following specific assets: solid house, refrigerator, vehicle, and phone respectively. All regressions include state dummies. Heteroskedasticity robust standard errors are presented in parenthesis. * significant at 10% ** significant at 5% *** significant at 1%.

B. Data

B1 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.[32] 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.[33] We use our assignment variable – the maximum groundwater depth recorded at any point over a three year period covering 2011-2013 – as our measure for well depth.[34] 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 A4.

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.[35] To provide some context, one standard deviation is approximately equivalent to 103 $litres/ha/day$.

[32] 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`

[33] 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.

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

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

Table A4: Constants used in water input calculation

Variable	Value	Units	Source
c	3.6×10^6		Ryan and Sudarshan (2020)
E	0.25		Ryan and Sudarshan (2020)
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 (2020) from case studies on irrigation pumping technology in India.

B2 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 Normalised Difference Vegetation Index (NDVI) and 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. Evidence suggests that NDVI values obtained from MODIS predict land use – to the extent of classifying general crop types – with 90% accuracy (Wardlow and Egbert, 2010). With respect to crop yields, Labus et al. (2002) find that NDVI values from MODIS are able to successfully predict the growth profile of wheat at both the regional and farm level in Montana, USA. In a study of millet in Senegal, Rasmussen (1997) estimates a correlation coefficient of 72% between NDVI and actual yield. Additionally, the author finds that subtracting early-season NDVI from the peak integral NDVI values significantly improves the level of explained yield variance – an approach we therefore adopt in our empirical analysis. Comparisons of EVI and NDVI have found that both indices produce equivalent crop classification results (Wardlow and Egbert, 2010) and are equally successful at predicting yields (Kouadio et al., 2014). EVI has been found to be especially sensitive to high biomass locations and tends to present relatively lower ranges over lower biomass sites (Gao et al., 2000).

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. Formally, NDVI is calculated as:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \tag{12}$$

where ρ_{NIR} (846–885 nm) and ρ_{RED} (600–680 nm) are the surface reflectance for the respective MODIS bands. 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.[36] The EVI is estimated in the same manner as the NDVI, but uses additional wavelengths from the blue band so as to account for atmospheric disturbance and background corrections. For an in-depth review of the literature and methods on calculating vegetation indices based on satellite imagery, refer to Huete et al. (2002).

As part of their research evaluating India's national rural road expansion programme, Asher and Novosad (2020) compiled data on the NDVI and EVI at the village-level across India 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 and a range of proxies for agricultural production – based on evidence from previous research evaluating the accuracy of vegetation indices from satellite imagery – were calculated for each year and season. This data is made publicly available as part of the replication material of their published paper.[37] We employ three proxies calculated for both the NDVI and EVI in each growing season: (i) the mean value (Mkhabela et al., 2005), (ii) the maximum value (Labus et al., 2002), and (iii) the difference between the maximum value and the early season value (taken as the average of the first three 16 day periods) (Rasmussen, 1997). This third proxy enables us to subtract interference from non-crop vegetation such as forest cover, thereby providing a more accurate measure for agricultural production. All our proxies are calculated as an average over a three-year period covering 2011-2013,

[36] 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`

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

coinciding with the time horizon of our assignment variable. Finally, for more interpretable results of our regression analysis, 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. Table A5 presents the replication of these results. Specifically, the authors ran panel regressions (2000-2006) of the differenced NDVI and EVI proxy 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.

Table A5: Correlates of NDVI and EVI proxies on district agricultural output

	Differenced NDVI (*ln*)		Differenced EVI (*ln*)	
	(1)	(2)	(3)	(4)
Output (*ln*)	0.331	0.233	0.235	0.197
	(0.042)	(0.040)	(0.046)	(0.041)
R^2	0.74	0.78	0.85	0.89
N	2124	2124	2124	2124
District FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes

Notes: This table replicates the results from Asher and Novosad (2020) evaluating the validity of vegetation indices as proxies of agricultural production. The NDVI and EVI proxy are based on satellite images, calculated as the difference between early season and the maximum value. Agricultural output is obtained from the Planning Commission's district-wise domestic product data. Heteroskedasticity robust standard errors are presented in parenthesis.

B3 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.

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.[38] 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.[39] Following the methodology of Elbers et al. (2003), Asher et al. (2021b) 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.[40] 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.[41] 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.[42] 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. (2021b) 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

[38] 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.

[39] 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`

[40] 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.

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

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

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 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.

B4 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. (2021b) 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. (2021b) 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.[43] 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.[44] We make use of data on the average pixel luminosity at the village level.[45] Specifically, we measure the average of mean night light for a village over a five year period, from 2009 to 2013, as our main proxy for economic activity (for the purpose

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

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

[45] Average luminosity in a given year is calculated by dividing total luminosity by the village area.

of the regression, this variable is log transformed).