Irrigation and the Spatial Pattern of Local Economic Development in India*

David Blakeslee† Aaditya Dar‡ Ram Fishman§

Samreen Malik¶ Heitor Pelegrina‖ Karan Singh**

December 2021

Abstract

We study the impacts of 1,500 large-scale irrigation projects that have affected more than 250,000 villages in India. To do so, we use high-resolution spatial data, and exploit discontinuities in program inclusion arising at project boundaries. Irrigation increases agricultural output and population density in rural villages. However, in and near towns, it causes a decline in indicators of development including population density, nightlight density, built-up area, and firm employment, reallocating productive factors away from non-agricultural activities. Results are consistent with a model in which the local effects of agricultural productivity gains depend substantially on the spatial organization of the economy.

Keywords: Irrigation, Urbanization, Development. *JEL codes:* 013, 014, 015, Q15, Q25, Q32

*We are grateful to seminar and conference participants at the Delhi School of Economics, Indian School of Business, Indian Statistical Institute, New York University Abu Dhabi, University of Maryland Baltimore County, University of Virginia and the 'Dams, Development and Water' conference co-organized by the Institute of Economic Growth, Delhi and University of Manchester, UK for helpful discussion and feedback. We would also like to thank Vedarshi Shastry and Chanchal Kumar Singh for excellent research assistance.

†Email: david.blakeslee@nyu.edu. New York University Abu Dhabi.

‡Email: aaditya_dar@isb.edu. Indian School of Business.

§Email: ramf@post.tau.ac.il. Tel Aviv University.

¶Email: samreen.malik@nyu.edu. New York University Abu Dhabi.

‖Email: heitor.pelegrina@nyu.edu. New York University Abu Dhabi.

**Email: vbk.singh@gmail.com. Independent researcher.

1 Introduction

Policy makers in developing countries have long emphasized improvements in agricultural productivity as a central strategy for promoting rural development. Ultimately, however, economic development hinges upon firm creation and shifting employment from the agricultural to the manufacturing and service sectors, a process often linked to urbanization and migration (Johnston and Mellor, 1961; Lewis, 1954; Gollin et al., 2002; Rostow, 1960; Kuznets, 1961; Studwell, 2013). It is, therefore, crucial to understand how gains in agricultural productivity impact non-agricultural development.

This paper studies the effects of permanent agricultural productivity shocks on local economic development in India. Since 1950, the Indian government has extended irrigation to close to 250,000 villages through the construction of large-scale dams and networks of canals that distribute river water to downstream villages. We provide evidence that these irrigation projects had, relative to non-irrigated regions, a positive impact on the agricultural productivity of rural villages by allowing them to expand crop production to seasons when it had previously been nonviable. We also show that the process of urbanization and the reallocation of labor to non-agricultural firm employment has slowed down within project areas.

Since the seminal work of Duflo and Pande (2007), a handful of papers have studied the impacts of surface irrigation projects on downstream areas, generally relying on exogenous variation in the geographical determinants of dam location for causal identification. These studies have documented important effects of irrigation on agricultural output, income volatility, and poverty rates (Hansen et al., 2011; Strobl and Strobl, 2011; Blanc and Strobl, 2014; Olmstead and Sigman, 2015; Jones et al., 2019; Dillon and Fishman, 2019; Zaveri et al., 2020). [1] However, less is known about the effects of irrigation on non-agricultural economic activity, which is the primary focus of our paper.

Our analysis uses fine spatial data on more than 1,500 major surface irrigation projects in In-

[1]Additional papers, including Hornbeck and Keskin (2014, 2015); Sekhri (2014); Fishman et al. (2013); Blakeslee et al. (2020); Ryan and Sudarshan (2020), have studied the impacts of decentralized groundwater irrigation on similar outcomes.

dia, which we merge with administrative village-level agricultural, demographic, and economic data, as well as remotely sensed land-use data. The boundaries of the areas served by these irrigation projects (also called "command areas") are primarily determined by engineering considerations related to topography (see Section 4 for details). We exploit the discontinuity in program inclusion arising at the boundary of command areas, comparing villages proximate to one another but on opposite sides of the boundary, while controlling for geographic features and imposing sampling restrictions to ensure comparability. This approach differs from much of the existing literature, which generally compares larger areas downstream from a dam to areas that are not.

The high resolution of our data allows us to provide novel insights on the impacts of irrigation on the spatial patterns of economic activity. A stylized fact about the nature of economic development in India is that village population density and non-agricultural economic activity decline rapidly with distance from rural towns. Our analysis shows that while the positive impacts of irrigation on agriculture productivity are broadly uniform, non-agricultural impacts display remarkable variation across this distance gradient. Far from towns, where nearly all labor is engaged in farming, we find no significant effect of irrigation on village labor force composition and a modest increase in small firm employment. In stark contrast, in towns (and also in their nearby villages), irrigation causes a substantial increase in the share of agricultural workers and a large decrease in employment in firms, especially large ones. Similarly, most villages experience an increase in population density, built-up area, and nightlight density as a result of irrigation while towns experience a decrease. Proxy measures of per-capita wealth are also higher in irrigated villages, while in towns there is generally no improvement.

To guide the interpretation of our empirical results, we formulate a parsimonious spatial economy model in which non-agricultural activities are subject to dynamic external economies of scale, as in Matsuyama (1992). In our model, a positive agricultural productivity shock in rural villages further deepens their specialization in agriculture and reduces the outflow of workers. In contrast, an identical shock to urbanizing locations raises the costs of labor and land, reducing non-agricultural activity and growth in manufacturing productivity, and slowing the inflow of workers.

Our paper joins a growing literature on the causal impact of different forms of agricultural productivity gains on structural transformation. McArthur and McCord (2017) and Gollin et al. (2018) utilize exogenous cross-country variation in the accessibility of key inputs to show that agricultural productivity gains accelerate structural transformation at the country level. Other papers use within-country variation to study the more local impacts of increases in agricultural productivity. Bustos et al. (2016, 2020) find that Brazilian municipalities endowed with favorable agro-climatic conditions for the adoption of labor-augmenting technical change experienced higher rates of local structural transformation. Hornbeck and Keskin (2015) show that U.S. counties more likely to gain access to irrigation after World War II experienced long-run improvements in agricultural output, but did not experience long-term increases in non-agricultural activity. Foster and Rosenzweig (2004), in the most similar context to ours, show that high rates of crop yield growth in India are correlated with lower industrial growth across a nationally representative sample of villages. Relatedly, several papers study how climatic variation affects urbanization and labor allocation in Sub-Saharan Africa and India, presumably through its effect on agricultural productivity (Henderson et al., 2017; Emerick, 2018; Krishnaswamy, 2019; Colmer, 2021).[2]

We contribute to this literature by providing causally interpretable estimates of the local impacts of agricultural productivity on long-term indicators of economic development and specialization. Moreover, the program studied covers hundred of thousands of villages, and occurs in a context where extending irrigation access has long been a key strategy for promoting rural development. We also examine the spatial patterns of these impacts at a finer resolution than previous papers, which generally conduct their analyses at higher levels of administrative aggregation that include both urban locations and their rural hinterlands. Though such analyses have provided extremely important insights about local impacts at these more aggregated levels, they are unable to shed light on the local spatial distribution of these impacts, or how they depend on baseline levels of urbanization or proximity to urbanized areas.

[2]See Barrett et al. (2017) for a review of studies in Africa.

Our methodology is less suited to identifying the aggregate regional or national impacts of increases in agricultural productivity, for which a structural model would be necessary for generating counterfactuals. Our estimated treatment effects are best interpreted as the local, marginal impacts of differences in agricultural productivity that take place against the backdrop of the economy-wide effects of irrigation expansion, as we discuss in detail below.

This paper also speaks to research on economic geography. Several papers have studied different drivers of the spatial distribution of economic activity in individual countries (e.g., Michaels et al., 2012; Bleakley and Lin, 2012; Allen and Donaldson, 2018; Davis and Weinstein, 2002). Most of this research has focused on developed countries. We contribute to this literature by providing evidence on the impact of agricultural productivity shocks on the spatial organization of production in India. By showing reduced-form evidence on how agricultural productivity shocks interact with the spatial distribution of economic activity, we complement recent papers studying interactions between structural transformation and economic geography, such as Gollin and Rogerson (2014), Nagy (2020), Eckert et al. (2018), Fajgelbaum and Redding (2018), and Henderson et al. (2018).

Lastly, our results relate to a rich literature studying the effects of agricultural productivity gains on various outcomes, including several papers on the "green revolution" (Christiaensen and Martin, 2018; Gollin et al., 2018; Bharadwaj et al., 2020; von der Goltz et al., 2020) and a small but burgeoning literature on irrigation (Hornbeck and Keskin, 2014; Dar, 2019; Blakeslee et al., 2020). Two concurrent papers that also examine the impact of irrigation in India are worth highlighting. Boudot-Reddy and Butler (2021) examine the impact of groundwater (well) irrigation and find that it increases agricultural production and consumption, but does not re-allocate labor across sectors.[3] Asher et al. (2021) study the impacts of canal irrigation using a similar research design to ours and find positive impacts on agricultural productivity and population density but no significant effect on overall structural change; and argue that irrigation may have driven labor

[3]The authors use a fuzzy regression kink design that exploits a technological constraint on the operational capacity of irrigation pumps.

movement across larger spatial scales.[4] While both papers derive results for aggregate population levels that are consistent with those in this paper, we show that these aggregate impacts may mask important reallocations of labor and economic activity towards villages and away from towns. In addition, we find that forms of economic activity characteristic of towns—in particular, manufacturing employment and large firms (≥50 workers)—suffer substantial declines that are not recovered by increases in villages.

2 A Simple Model of Agricultural Productivity and Specialization

This section develops a spatial economy model, in which productivity in urban activities are subject to dynamic external economies of scale as in Matsuyama (1992). Our model predicts theoretically heterogeneous effects of an agricultural productivity shock on non-agricultural activities, depending on the distance of affected regions to population hubs.[5] We begin with a simplified economy with one town and one village. We let workers move between these two sub-regions, but not to other parts of the economy. This allows us to more clearly highlight the mechanisms generating the heterogeneous effects of agricultural productivity shocks. In the end of this section, we discuss extensions with labor mobility between multiple villages and towns and the existence of a non-tradable sector.

Setup. Consider a small open economy with two sub-regions, called Town (T) and Village (V), and two sectors, agriculture (A) and manufacturing (M). The economy operates over discrete time. Relative to Village, Town is situated in a privileged geographic area, near a major trade route or a river, such that trade between Village and the rest of the world (ROW) has to pass through Town. To take goods from Village to Town, there is an iceberg trade cost of τ. This

[4]The authors implement a regression discontinuity design, using elevation relative to a canal as the running variable. Both research projects have been developed independently and in parallel.

[5]We relegate details and derivations to Appendix Section A1.

economy has a population N that can move between regions and sectors. Each region has a land endowment of L. Markets are perfectly competitive. Workers are perfectly mobile within districts, but not between.

Technologies are given by

$$q_{ikt} = A_{ikt} \left(L_{ikt}\right)^{\alpha_k} \left(N_{ikt}\right)^{1-\alpha_k}$$

where i indexes a region, k a sector, and t a time. A_{ikt} is the productivity, L_{ikt} the employment of land, N_{ikt} the total labor employment, and α_k the cost share of land. Agriculture is land-intensive ($\alpha_A > \alpha_M$). Agricultural productivity is fixed ($A_{iAt} = A_{iA}$). As in Matsuyama (1992), manufacturing productivity is subject to knowledge accumulation:

$$A_{iMt+1} = A_{iMt} + \gamma n_{iMt}$$

where $\gamma > 0$ is an externality parameter and n_{iMt} is the share of workers in manufacturing ($n_{iMt} \equiv N_{iMt}/N_{it}$, where N_{it} is the total population in i). We assume $A_{TA} = A_{VA}$ and $A_{TM0} > A_{VM0}$, where $t = 0$ is the initial period.

In every region, land consists of a continuum of plots and landowners assign plots to sectors. Each plot requires a conversion cost of ϵ_k to be employed in sector k, incurred by the landowner. Similar to Sotelo (2020), this conversion cost is heterogeneous across plots and drawn from a Fréchet distribution, $F\left(\epsilon\right) = 1 - exp(-\epsilon^{-\theta})$, where θ controls the variance of conversion costs. With this formulation, as they maximize profits, landowners assign a share $\lambda_{ikt} = r_{ik}^{\theta} / \sum_{k'} r_{ik'}^{\theta}$ of plots to sector k, where r_{ik} is the rent of converted land.[6]

Lastly, consumers have Cobb-Douglas preferences.[7] The expenditure share of agriculture and

[6] Land heterogeneity avoids non-degenerate equilibria in which a region fully specializes in one of the sectors. We notice that, in contrast to Sotelo (2020), who model plots of land as heterogeneous in terms of their productivity, we assume that plots of land, once converted to a sector, are homogeneous. Our approach simplifies the solution of the model while retaining our goal of avoiding full sectoral specialization.

[7] We assume homothetic preferences to simplify the exposition. Our results would remain if we assume, for example, Stone-Geary preferences with a subsistence consumption parameter.

manufacturing are, respectively, μ_A and μ_M.

Dynamics and the Spatial Distribution of the Economy. The evolution of prices in Town is exogenous, tracking prices in the ROW.[8] In Village, prices are determined by non-arbitrage conditions across space. Because Town is closer to the ROW, it tends to attract workers and have a larger population density. As such, land rent to wage ratios in Town are higher, inducing specialization of Town in manufacturing and specialization of Village in agriculture. In any given period, both population density and the share of manufacturing workers are larger in Town.

Over time, productivity growth in Town is larger than in Village because of knowledge accumulation in manufacturing. As a result, Town increasingly attracts workers and specializes in manufacturing. In Village, population density falls over time, and the region increasingly specializes in agriculture.

Agricultural Productivity Shocks. Consider now a shock at $t = 0$ that increases agricultural productivity permanently, either in Town or in Village. In both cases, this shock reduces permanently the local share of workers and land in manufacturing relative to a scenario without the shock. In Town, this reduction slows down the productivity growth in manufacturing, reducing the inflow of workers over time. In Village, this shock also reduces the productivity growth in manufacturing, but it tends to prevent the outflow of workers by making the agricultural sector more attractive, increasing population in Village relative to a scenario without the shock. (Appendix Figure A1 illustrates these effects using numerical simulations of the model.) Of course, if baseline labor allocation is already entirely dominated by agriculture, the effect of the shock on labor reallocation can be muted.

In summary, our model predicts that an agricultural productivity shock increases the share of workers in agriculture, both in Village and in Town, which is consistent with empirical findings in Foster and Rosenzweig (2004). However, our model also predicts starkly different effects on population growth, depending on whether the shock affects a village or a town. This last

[8] Appendix A1 characterizes prices in the ROW.

prediction is consistent with the new empirical findings that we present in Section 5.

Extensions. We briefly consider several potential extensions to the model. Let us start with the existence of a local non-tradable sector that produces goods and services using labor. In this extension, the agricultural productivity gain in Village would increase the local demand for services in Village and therefore expand the labor employment in non-tradables. In Town, this shock would instead reduce the expansion of the non-tradable sector, since in-migration from other regions would slow down.

Next, consider a model with multiple villages and towns, between which workers can move with some distance-dependent migration cost. In this case, an agricultural productivity gain in a particular village will not only prevent the out-migration of workers to a nearby town, but may also attract workers from other villages, which would reinforce the local impact of the agricultural productivity shock in our model.

Finally, we consider a model which again has multiple villages and towns, but in which there is also a service sector that is tradable (only) across geographically proximate villages and towns. In this case, an agricultural productivity gain for a particular village may lead to the retention of workers both in that village as well as in other nearby villages, because the affected village will increase its demand for services from nearby areas.

These considerations will affect the interpretation of our empirical findings, as we discuss in greater detail in Section 4.

3 Data

We make use of a variety of data sources available at high spatial resolution. The key outcome variables come from: (a) demographic and economic censuses, available at the village and town level; and (b) remotely sensed data on cropping patterns, land use, and nighttime lights. The latter are merged to georeferenced villages and towns, along with GIS data on canal command areas and key geographic factors. Additional details are provided in Appendix A2.

Demographic and economic census. The demographic census of India is conducted every ten years. It includes data on demographics, economic activity, educational attainment, land use patterns, and household amenities and assets for the entire country, aggregated at the village and town level. We make use of the following outcomes from the 2011 census: irrigated area, canal-irrigated area, population density (per sq km), labor force participation, employment in agriculture (both own-farm cultivators and agricultural laborers), and ownership of assets and household amenities. We also use data from the sixth edition (2012-13) of the economic census, which provides firm-level data on employment for all enterprises in the country, including both the sector and number of workers within each firm. It is important to note that, while the demographic census reports the numbers of workers and farmers residing in the village, the economic census reports the number of employees of firms which are located in the village/town, whether they reside in it or not.

Remotely Sensed Data. We use three sources of satellite data with information on agricultural outcomes. First, we utilize data on dry season cropping from MODIS Enhanced Vegetation Index (EVI) to measure cropped area at small-scale farming environment (Jain et al., 2017). The data are available at a 1×1 sq. km resolution, and aggregated using village and town polygons. Second, we use land use and land cover classification (250K) data from *Bhuvan*, the Indian Space Research Organisation's (ISRO) online portal.[9] The data are made available by the Natural Resources Census programme at National Remote Sensing Centre (NRSC), which uses remote sensing to estimate land use in different categories, including: season-wise cropping, double or triple-cropping, fallow area, built-up area, forest area, wasteland, and water bodies. These data are used to estimate net sown area in the country, as they have a high accuracy (Agency, 2007). Third, as a proxy for economic growth and urbanization, we use nighttime lights data from NOAA's National Geophysical Data Center's Defense Meteorological Satellite Program (Henderson et al., 2012). The extensive use of remotely sensed data in this paper, including novel data from Indian satellites, is used to complement the analysis from administrative data which might be prone to measurement

[9] http://bhuvan.nrsc.gov.in/gis/thematic/index.php

error (Donaldson and Storeygard, 2016).

Spatially Linked Data. Using village and towns polygons, we combine the data sets described above to construct a high resolution spatial data set on economic activity in the country. We also merge GIS data on canals, command areas, aquifers, and rivers from the India Water Resources Information System (WRIS).[10] Attribute data on canals is completed using Central Water Commissions' Management Information System of Water Resources Projects and India WRIS Wiki.[11] Finally, we calculate distances from village centroids to command area boundaries, and complement the data with detailed information on geographical features including climate, altitude, slope and a land ruggedness index formulated by Riley et al. (1999), and used by Nunn and Puga (2012) and Michaels and Rauch (2017).

Summary Statistics. Appendix Table A1 reports the sample size and descriptive statistics. The sample cover approximately 1,500 irrigation projects; and includes approximately 74,000 villages and 900 towns within program areas, and similar numbers in nearby control areas. To put these numbers in perspective, there are approximately 650,000 villages and 7,700 towns in India. Therefore, our sample of treated villages and towns accounts for approximately 11-12 percent of all villages and towns in India.[12]

Descriptive statistics for villages are given in column (1), and columns (2)–(4) report the differences between towns and villages, with column (3) including command area fixed effects, and column (4) additionally restricting the sample to towns smaller than 30 sq kms. As is apparent, towns are systematically different than villages, having larger populations, smaller agricultural sectors, more large firms (per capita), and greater household asset holdings. These significant differences across villages and towns also motivate the spatial analysis and the empirical strategy

[10]Data downloaded from http://59.179.19.250/ during Nov 2019–Apr 2020. The link, however, is now inaccessible.

[11]https://indiawris.gov.in/wiki/doku.php

[12]Out of the 567,125 villages for which data is available, 16 percent do not get any irrigation. Overall, the average percentage of cultivated land in Indian villages that is irrigated from any source is 52 percent and the average cultivated area irrigated by canals is 12 percent. After tubewells, canal irrigation is the second most popular means of irrigation and close to 144,000 villages (24 percent of all villages in the country) report receiving water from canals.

we describe next.

4 Empirical Strategy

Our empirical strategy exploits the discontinuity in program inclusion arising at the boundary of command areas, comparing villages (towns) proximate to one another on opposite sides of the boundary. Command areas are defined as the total areas to which an irrigation project can deliver water through a network of canals. The extent of the command area is determined by the volume of water in storage (mostly in a dammed reservoir, but occasionally through direct diversion of an un-dammed river) and the topography of the terrain. Since water is distributed through gravity, elevation plays a key role in determining the boundary. In one of the most common engineering designs, the main canals begin at the dam and follow a roughly constant elevation contour, from which secondary canals deliver water to lower elevations. The command area boundary is thus formed by these main canals. In another common design, the main canals follow ridge lines and secondary canals distribute water to both sides of the ridge. The boundary of the command area is then defined by lowest elevation lines on both sides of the ridge and the terminus of the main canals. Using elevation data, we confirm that the the command area boundaries are essentially flat, with average slopes on the order of a 20 cm decline per 100 meters distance.

Our analysis encompasses approximately 1,500 irrigation projects (i.e., command areas), for which we have high resolution data on the boundaries of command areas, as well as canals and all relevant geographic features. To improve the comparability of the control and treatment groups, we restrict the sample to villages and towns whose centroids are no farther than 10 kms from the boundary (see Figure 1), but our results are not affected by the choice of a narrower or wider bandwidth.[13] To sharpen the comparison even further, we partition the boundaries of command areas into 5 km segments (of which there are, on average, 16 per command area), and compare

[13] Given that there is no well accepted method to select bandwidth in a multi-dimensional regression discontinuity (Dell and Olken, 2020), our chosen bandwidth is one of the most conservative in the literature in comparable contexts. Prior border design studies set in a developing country typically have a bandwidth between 25 km and 200 km (such as, Dell, 2010; Michalopoulos and Papaioannou, 2013; Dell and Querubin, 2018).

only villages on opposite sides of the same boundary segment.

Formally, our main estimation takes the form:

$$y_{ipdb} = \alpha + \beta C_i + \mathbf{X_i}\Gamma + \nu_d + \mu_b + \varepsilon_i, \tag{1}$$

Agricultural and non-agricultural outcomes are denoted by y_{ipdb}, where i is an index for location (village or town) in a 10 km buffer around irrigation project p in district d and b is an index for 5 km command area boundary segments. The key explanatory variable of interest, C_i, is a binary variable indicating whether the centroid of the location lies within a command area or not, and the coefficient of interest is β which is the impact of irrigation on agricultural productivity and local economic development. Our preferred specification includes district fixed effects, ν_d, and μ_b which are the 5 km boundary segment fixed effects, assigned according to the boundary segment to which a centroid of a given location is closest.[14] We also control for a vector of village geographic characteristics, X_i, which includes altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the village, and the (log) area of the village. We discuss these in detail below. To account for spatial correlation, error terms are clustered at the command area level.

We subject our results to several robustness tests. First, we consider alternative choices of bandwidths and document that the results in both villages and towns are robust to varying the bandwidth between 2 km and 30 km. Second, we control for (a linear spline in) the distance from the village to the command area boundary (omitting villages that are partially inside the command area), as is customary in spatial discontinuity designs carried out over larger spatial scales. It is important to note that the narrow extent of the spatial sample we use for our estimation makes such controls less crucial, while the possibility of spillovers undermines one of the key requirements of this research design. Third, we use Conley standard errors that account for spatial correlation across villages at distances of up to 300 kms. Fourth, we winsorize the outcome

[14] In the Appendix, we include additional analysis which replaces the boundary segment fixed effects with project fixed effects while maintaining district fixed effects.

variables at the 5th and 95th percentiles.

Our approach is similar in spirit to spatial regression discontinuity designs that have been employed in a number of papers (Dell, 2010; Sukhtankar, 2016; Dell and Querubin, 2018; Dell and Olken, 2020; Ahlfeldt et al., 2015; Egger and Lassmann, 2015; Gonzalez, 2021; Smith, 2019). The identifying assumption in such designs is that other than the treatment, all factors that can potentially affect the outcomes of interest vary smoothly at the boundary. In our case, this assumption is motivated by the plausible argument that prior to the construction of an irrigation project, there would be little reason to expect the command area boundary, determined as it is through a highly specific function of topography and the volume of the reservoir, to coincide with substantial breaks in other geographical or socio-economic variables. A similar argument is made by Jones et al. (2019) and Blakeslee et al. (2019), who evaluate specific surface irrigation projects in Rwanda and India, respectively.

Towns vs Villages We estimate treatment effects for villages and towns separately, as our model implies divergent impacts for these two categories of settlement. Township status is based on three factors: population level, population density, and the share of the non-agricultural labor force.[15] As seen in Appendix Table A1 and Appendix Figure A2, these features are correlated with more fundamental differences in the demographic and economic structure of a settlement. One potential concern with this approach is that whether a village graduates to the status of being a town is itself endogenous, which may confound the estimation of treatment effects with composition effects. Appendix Figure A7, which charts the number of towns by year (demeaned by the number in 1951) in the sample and out-of-sample areas, shows the substantial growth in the number of towns during this time. Most towns were formed after 1991, by which time the vast majority of irrigation schemes had been completed.

In Table A2 we test for endogenous town formation. In column (1), we restrict the sample to villages and towns that were close to meeting the criterion for township formation (the 'marginal

[15]Formally, the Census of India classifies a 'census town' as a settlement if the following three conditions are met: the population exceeds 5,000, the population density is more than 400 persons per sq km, and more than 75 percent of main male working population is employed outside the agricultural sector.

sample') and estimate the impact of being in the program areas on attaining township status.[16] In columns (2) and (3), we restrict the sample to all towns, and take as the outcome variable an indicator for whether the town already had township status in 1951 and 1971, respectively. In columns (4)–(5), we take as the outcome the log *area* within towns, where land within villages take a value of 0, and towns take the natural log of their areas.

We find no impact on whether villages graduated to township status, nor on how early existing towns were formed. In addition, we find no impact on the total area of the command area that is within a township (column 4), or on the size of towns (column 5). Essentially, this means that the (subsequently estimated) 6.3% increase in village population in treatment areas—coupled with the absence of any change in the labor share in agriculture—was insufficient to graduate villages to township status. This is intuitive, given the dramatically larger populations of towns, and their far smaller agricultural labor shares. It is important to reiterate that this does not mean that irrigation had no effect on the rate of town formation in the aggregate or over larger spatial scale. Rather, such a response does not seem to occur near the project borders, which allows us to separate our causal estimates between towns and village.

Interpretation and relation to the model In interpreting the relation of our empirical estimates to the model presented in Section 2, several remarks are in order.

First, it is important to differentiate between the economy-wide and the local effects of irrigation. The large-scale introduction of irrigation is likely to generate substantial economy-wide impacts, including the aggregate rate of structural transformation and the spatial allocation of labor. The identification strategy employed in this paper, however, is unable to shed light on these non-local effects, which ultimately require structural estimation. Similar to previous related papers, such as Foster and Rosenzweig (2004) and Bustos et al. (2016), we avoid making claims about the direction and magnitude of these economy-wide impacts, and focus on effects that are local and amenable to causal inference. The impacts we estimate should therefore be interpreted as

[16]We define a 'marginal sample' as villages and towns that were close to meeting the criterion: specifically, those with a population between 4,000 and 6,000 people, a population density of more than 350 persons per km sq, and male labor force greater than 70 percent that is engaged in non-agricultural production.

the local, long-term effects of increases in agricultural land productivity, which occur against a benchmark of these possible economy-wide impacts of irrigation projects.[17]

Second, while the model discusses the impacts of a particular village or town receiving a positive agricultural productivity shock, our empirical analysis is based on regional irrigation projects that affect multiple contiguous settlements. The RD approach seeks to isolate the causal impact of irrigation on a particular village (or town) by comparing it to a nearby village (or town) situated on the opposite sides of the boundary, which would otherwise be plausibly statistically identical.

However, if labor movements are strongly constrained by distance—so that migration primarily occurs across geographically proximate locations—then this has important implications for the interpretation of our results. In treated towns, the slowdown in in-migration may be the result both of lower non-agricultural productivity; as well as higher agricultural productivity in nearby (treated) villages, which increases agricultural wages and reduces out-migration to nearby towns. In treated villages, the increase in population density may derive not only from slower out-migration to towns, but also from in-migration from nearby untreated villages on the other side of the boundary, which would bias the results.[18] As we shall see later, we find evidence consistent with a reduction in population movements *within* treatment areas from villages to towns, but little evidence for population movements between villages *across* the control and treatment areas.

Threats to Identification We consider two principal threats to the identifying assumption that the control represents a valid counterfactual to the treatment. The first relates to potential differences in geography across the command area boundary, which may arise if engineering considerations result in command area boundaries that coincide with breaks in certain geographical

[17] Asher et al. (2021), in contrast, raise the possibility that irrigation may have increased town formation in the broader regions surrounding command areas, but add the caveat that the estimates underlying this conclusion are not well identified.

[18] If migration were not constrained by distance, then any decline in migration from treated villages to treated towns would be off-set by the migration to treated towns from more distant (out-of-sample) areas. Similarly, absent migration frictions, any in-migration to treatment-area villages would be primarily due to migration from more distant areas rather than nearby control villages.

features of the terrain. For example, it may be deemed optimal to place the boundary along the base of a hill or the border of a forested area.

Figure 2 displays plots (black lines) of key geographic variables (altitude, type of aquifer, ruggedness, and distance to river) against the distance between a village and the nearest command area boundary. The plots do not indicate discontinuous jumps, but do suggest trend breaks in elevation and ruggedness. However, when we limit the sample to villages lying in the vicinity of boundary segments for which the average slope on both sides is very moderate (less than 1.5 degrees), no such trend breaks are visible (blue lines in Figure 2). We therefore use this sampling restriction in our analysis.[19] Appendix Figure A3 displays the geographic coverage of the trimmed sample.

Because the geographic variables generally trend monotonically with elevation, and because the latter is one of the key determinants of inclusion in the program area, small differences in geographic characteristics will necessarily be present across the boundary even under our conservative sampling restriction. For this reason, we control for all of these variables in our regressions. In practice, however, the magnitude of the differences is small and of negligible agricultural significance (Table A3).[20]

The second threat to identification is posed by the possibility that non-engineering considerations may influence the boundaries of the irrigation project, such as the desire to include politically favored villages in the command areas. If differences in outcomes across the boundary were driven by unobservable factors associated with such favored villages, one would expect treatment effects to be particularly large at the boundary, and to decline at greater distances. As we show below, we find no evidence for such patterns in plots of outcomes against distance to the boundary, nor do we find materially different treatment effects when omitting villages just inside the command area from our regressions.

[19] Canal boundaries where the canal is within 500m of a river are also excluded.

[20] For example, there is a 5 meter elevation difference between control and treatment villages (10 km bands), in comparison to a control mean of 200 meter, amounting to 0.01 standard deviations. Ruggedness differs by only 2 points on the Riley index, compared to a control mean of 39, where any value of this index between 0 and 80 is considered level terrain.

Several additional tests of the identification assumption are reported in the results section. This includes a placebo analysis using only those projects that were initiated after the year 1991, and testing whether treatment effects are apparent for 1991 outcomes (using the same regression specification). In addition, we conduct an analysis limiting the sample to only those boundary segments that are demarcated by irrigation canals. Because such canals follow approximately fixed elevation contours, and the command area consists exactly of the area on their downhill side, treatment status for villages along these segments is determined by transparent and fundamental engineering considerations.

5 Results

Agricultural Outcomes. In our first set of results, we present the impact of being included in the command area on agricultural outcomes, including: the percentage of agriculture land that is irrigated; the share of land that is used for multiple-season cropping; and the extent of dry season cultivation (EVI). We show that the effects on agricultural activity are substantial and similar between rural villages and towns.

We illustrate the results graphically in Figure 3.1–3.3, which plots these outcomes (or rather, residuals from regressions of these outcomes on all control variables in specification 1) against distance bins from the boundary, labelling distance as negative within the command area and positive outside of it.[21] Results for regressions without controls are depicted in Appendix Figure A5.1–A5.3. All three outcomes display clear discontinuities at the boundary.

We report regression estimates using alternative agricultural outcomes in Tables 1 and 2. Within command areas, the share of agricultural land that is irrigated by canals increases by around 8.4 percentage-points (p.p.), representing a more than 150% increase over the control mean (5.1 p.p.).[22] These effects are large in proportional terms but modest in magnitude, consistent with the generally poor assessment voiced by observers of the success of these projects in increasing

[21] The plot excludes villages which overlap the boundary and for which treatment status is poorly measured. These villages are typically located within 2-3 kms of the boundary.

[22] Census data on irrigated and cultivated areas are only reported for villages.

irrigated area. Canals are one of several potential sources of irrigation raising the possibility that substitution to other sources may attenuate the net effect on irrigation. However, the overall share of irrigated agricultural area increases by 5.6 p.p., representing a 13% increase over the mean value outside the command area. We also estimate a 7.0 p.p. increase in the remotely sensed share of cultivated village area, a 7.3 p.p. increase in the share of land with multi-season cropping, and an increase in dry season vegetation indices (EVI) (Table 2).

The estimated effects in towns are somewhat larger—except for vegetation indices, which are smaller and imprecise—but they are not statistically different from the effects on villages. Though we lack data on agricultural yields at the required spatial resolution, the clear discontinuities in these outcomes at the boundary and the increase in the number of crops grown in a single year suggest a substantial increase in *annual* agricultural output per acre.

Consistent with our theoretical analysis in Section 2, the estimated impact of canals is not substantially different between towns and villages. In both types of regions, there is a similar increase in measures of agricultural activity, as captured by remotely-sensed data. Next, we turn to the impacts on urbanization and development, where we instead find that the impact of canals on non-agricultural activities is substantially different between towns and villages.

Urbanization and Development. This section presents the impacts of canals on urbanization and development, which we measure through the distribution of population, built-up area, and nightlight density. In particular, we show that the effects on measures of development are substantially different between rural villages and towns.

Similar to the illustration for agriculture outcomes, we present our results for urbanization and development for villages graphically in Figure 3.4–3.6. Results for regressions without controls are depicted in Appendix Figure A5.4–A5.6. All three outcomes display clear discontinuities at the boundary.

In Figure 4.1 and Table 3 we report estimates of the impact of canal irrigation on these outcomes (measured in logs) for villages and towns separately. For villages, we estimate a 6.1%

increase in village population density, a 6.5% increase in light density, and a 3.5% increase in the built-up area. For towns, however, we observe opposite effects, with a 30.8% decline in population density, a 26.1% decline in light density,[23] and a 26.8% decline in built-up area. These opposing effects for villages and towns are consistent with the ambiguous impact of agricultural productivity shocks highlighted in our model. To appreciate the magnitude of these effects, it is worth benchmarking them against the modest (13%) effect on irrigated area, implying irrigation elasticities for these outcomes of substantial magnitudes.

Labor Force Composition. In Figure 4.2 and Table 4, we document the impact of canal irrigation on labor force participation and composition using demographic census data. We do not find significant effects in villages; but, in towns, we estimate an increase of 3.3 p.p. (24%) in the share of workers engaged in farming, driven by increases in both land-owning cultivators and landless agricultural laborers.

Firm Activity. We also examine impacts on firm activity, which we measure through the (log) employment in firms which are located in a given village or town, by sector and size. Results are depicted in Figure 4.3 and reported in Appendix Table 5 in greater detail. Employment in firms increases by 5.8% in villages, with effects evident for manufacturing (4.6%) and service firms (7.2%). The effects are mostly driven by small firms (less than 10 workers). For towns, in contrast, we find large, negative effects, with firm employment being 58.3% lower in command areas, which is driven by declines in both manufacturing (73.3%) and services (47.5%). Importantly, there are particularly large declines in all sizes of firms, where employment is more than 50% lower.

Assets. Figure 4.4 and Table 6 report estimated impacts of canal irrigation on various measures of asset holding and home amenities. In villages, we see substantial increases in the fraction of households owning most types of assets and the quality of housing facilities. In contrast, we find

[23] This effect is statistically insignificant when we include district fixed effects instead of 5-km boundary fixed effects.

no evidence for corresponding effects on asset holdings in towns.[24]

Additional Discussion of Identification and Robustness. We perform several additional estimations that provide indirect tests of our empirical approach. First, Appendix Table A4 and Appendix Figure A4 repeat the village estimation for key outcomes while restricting the sample to command area boundaries which are formed by irrigation canals. The results from this alternative identification strategy, which exploits plausibly exogenous variation stemming from fixed elevation contours (described in Section 4), remain similar.

Second, Appendix Table A5 presents a placebo analysis which limits the sample to villages for which the nearest command area was initiated *after* 1991, and outcomes are measured through the 1991 demographic census, 1993 light density, and 1990 economic census firm employment. We find no statistically significant impacts on any of the key outcome variables, and the point estimates are an order of magnitude smaller than in our main analysis, providing added confidence in our approach.

Third, we estimate our main results by controlling for (a linear spline in) the distance from the village to the command area boundary (omitting villages that are partially inside the command area). Appendix Table A6, Panel A presents estimates for villages while Panel B reports estimates for towns. A comparison of these results with those from equation 1 show that adding distance to boundary controls are less crucial as both the point estimates and statistical significance are very similar to the main results.

Fourth, while we present our results using the 10-km bandwidths, in Appendix Figures A6 we also use alternative bandwidths ranging from 2km–30km. We find that point-estimates are relatively stable across specifications.

Fifth, we ask whether the results are driven by the deliberate manipulation of the command area boundary to include certain favored villages. For this, we re-estimate impacts on key out-

[24]One mechanism through which assets become higher relative to non-irrigated areas in our current model is because land rents rise with agricultural productivity. Another way to rationalize the impact of irrigation on assets is to add imperfect mobility between regions. In that case, there would be a temporary gap between the real income of workers from different origins.

comes while removing the treated villages that are closest to the boundary (within 2 km). These are the villages which are most likely to be driving manipulation of the boundary. Were the treatment effects in fact being driven by unobservable attributes of these influential villages, then we would expect the treatment effects to decline with the exclusion of these villages. Reassuringly, the results are essentially unchanged both in magnitude and significance (Appendix Table A7).

Sixth, we estimate our main results while removing villages which intersect the boundaries (see Appendix Table A8, Panel A); with winsorized outcome variables at the 5th and 95th percentiles (Appendix Table A8, Panel B); and with Conley standard errors that account for potential spatial correlation in errors across villages that are up to 300 km apart (Appendix Table A8, Panel C). The results are not materially affected.

Lastly, we highlight that our identification strategy estimates relative effects at local geographic levels, where productive factors can reallocate between treatment and control areas. At this local level, our results indicate that we estimate lower-bound effects of irrigation: when we inspect the effects on light density and firm employment, for example, we find evidence of positive spillovers to control groups (see Figure 3.1).[25]

Heterogeneous Treatment Effects by Proximity to Towns. To better understand how agricultural productivity shocks interact with the spatial organization of the economy, we next explore whether treatment effects for villages vary by distance to towns. This analysis is motivated by Appendix Figure A2, which depicts a strong relationship between distance to the nearest town and a variety of demographic and economic variables (with distance set at 0 for towns themselves).

Table 10 reports similar regressions for household assets and home amenities. Results indicate that villages close to towns are somewhat worse off on a per-capita basis in command areas.

Figure 5 plots the magnitude of treatment effects for villages at various distances from the nearest town. The effects of irrigation on village population and built-up areas are positive further

[25]Our method does not capture, however, spillovers that have a more aggregate nature, such as migration to more distant regions or long-distant trade. Recovering such aggregate effects is not in the scope of this article.

from towns, but in their vicinity become negative: villages within 2 kms of a town experience an approximately 10% decline in population density (Figure 5.1) and built-up land (Figure 5.2). We also find that increases in the share of farmers in the workforce documented for towns also occurs for villages in the vicinity of towns (Figures 5.3; and that the same is true of employment in manufacturing firms 5.4). Tables 7, 8, and 9 present corresponding estimates and robustness tests, using a (treatment-interacted) binary indicator for town-proximity which takes a value of 1 for villages within 4 kms of the nearest town.

6 Conclusion

Over much of the 20th century, the construction of large-scale surface irrigation infrastructure was one of the most capital-intensive investments by governments wishing to boost agricultural economies in low and middle income countries. This paper evaluates the impacts of such irrigation projects in India, one of the countries which has pursued this strategy most vigorously since its independence.

Surface irrigation projects have long been criticized for their inefficient performance. While confirming the relatively modest local impact of these projects on irrigation, we nonetheless find important impacts on local patterns of economic development.

The impacts we find are sharply mediated by the proximity of villages to rural towns. In more distant villages, irrigation increases population density, night light density, and built-area, while also modestly increasing per-capita wealth. In towns themselves, population, nightlight density, and firm activity are reduced in irrigated areas, and greater shares of the labor force are retained in agriculture. Similar impacts are estimated for villages in the vicinity of these towns. These results are consistent with a simple spatial economy model in which the same permanent agricultural productivity gains can have substantially different results, depending on the geographic incidence of this shock.

This remarkable heterogeneity in impacts would be masked by estimates derived from the aggregate sample of settlements that includes both villages and towns. If anything, such aggregate estimations would need to account for the sharply uneven distribution of the population between villages and towns, but the endogeneity of the required population weights complicate such an exercise. Merely for illustration, we note that a back-of-the-envelope calculation that takes such weights into account suggests: no net difference in overall population between the inside and outside margins of the irrigation project borders; a 4.3% increase in agricultural workers; a 25% decline in employment in manufacturing firms; a 31% decline in employment at large firms ($\geq$50 workers); and a 3% increase in light density.

The ability to simultaneously conduct our analysis at a fine (village-level) spatial resolution and on a country-level scale allows us to estimate local impacts of surface irrigation that are both well-identified and externally valid. It is important to stress, however, the local nature of the treatment effects being captured by our identification strategy, and our inability to capture the economy-wide impacts of irrigation expansion on structural transformation and economic growth. These effects may cause a level increase in local economic activity across both control and treatment areas that is not captured through our identification strategy.

The conceptual framework that guides our analysis interprets the relative increase in population density in irrigated areas as reflecting a reduced propensity to migrate to adjacent towns. However, we are unable to empirically distinguish this and other potential mechanisms, such as an increase in fertility rates or the in-migration of workers from other villages in India, whether in the vicinity or not. We note, however, as reported above, that treatment-area village population increase closely mirrors the reduction in treatment-area town population, leading to an absence of an aggregate population effect, consistently with the interpretation in our model. In addition, in results not shown we find that the treatment effect for population density in 1991, before which migration rates were generally quite low (Lusome and Bhagat, 2006; Government of India, 2017), are virtually identical to those in 2011.[26]

[26] It should also be noted that town growth was if anything slightly *lower* in study sample than in out-of-sample areas (Appendix Figure A7), suggesting that any impacts of irrigation on *local* urban growth were relatively modest against the backdrop of wider urbanization patterns.

Overall, we find that local agricultural productivity gains arising from irrigation expansion can bring substantial benefits to rural farmers, but that they can also potentially hinder local non-agricultural economic activity in relatively more urbanized areas, consistent with findings by Foster and Rosenzweig (2004). We provide evidence that these agricultural productivity shocks have changed the spatial organization of agriculture, with potentially important implications to aggregate welfare. More complete investigations of the aggregate non-local effects of irrigation are left to future research, as they would require structural estimation, and are unlikely to be achievable through reduced form approaches employed in existing work.

References

AGENCY, N. R. S. (2007): "National Land Use and Land Cover Mapping Using Multi-Temporal AWiFS data," *Unpublished.*

AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

ALLEN, T. AND D. DONALDSON (2018): "The Geography of Path Dependence," *Unpublished manuscript.*

ASHER, S., A. CAMPION, D. GOLLIN, AND P. NOVOSAD (2021): "The Long-run Development Impacts of Agricultural Productivity Gains: Evidence from Irrigation Canals in India," *Working Paper.*

ASMAL, K. ET AL. (2000): "Dams and Development: a New Framework for Decision-making. The Report of the World Commission on Dams." *Dams and development: a new framework for decision-making. The report of the World Commission on dams.*

BARRETT, C. B., L. CHRISTIAENSEN, M. SHEAHAN, AND A. SHIMELES (2017): *On the Structural Transformation of Rural Africa*, The World Bank.

BHARADWAJ, P., J. FENSKE, N. KALA, AND R. A. MIRZA (2020): "The Green Revolution and Infant Mortality in India," *Journal of health economics*, 71, 102314.

BISWAS, A. K. AND C. TORTAJADA (2001): "Development and Large Dams: A Global Perspective," *International Journal of Water Resources Development*, 17, 9–21.

BLAKESLEE, D., R. FISHMAN, V. PATEL, AND Y. ROTHLER (2019): "Evaluating the Ramthal Irrigation Project: Short-Term Impacts," *Unpublished manuscript.*

BLAKESLEE, D., R. FISHMAN, AND V. SRINIVASAN (2020): "Way Down in the Hole: Adaptation to long-term water loss in rural India," *The American Economic Review*, 110, 200–224.

Blanc, E. and E. Strobl (2014): "Is Small Better? A Comparison of the Effect of Large and Small Dams on Cropland Productivity in South Africa," *The World Bank Economic Review*, 28, 545–576.

Bleakley, H. and J. Lin (2012): "Portage and Path Dependence," *The Quarterly Journal of Economics*, 127, 587–644.

Boudot-Reddy, C. and A. Butler (2021): "Watering the Seeds of the Rural Economy: Impact of Tube-Well Irrigation in India," .

Bustos, P., B. Caprettini, and J. Ponticelli (2016): "Agricultural Productivity and Structural Transformation: Evidence from Brazil," *The American Economic Review*, 106, 1320–65.

Bustos, P., G. Garber, and J. Ponticelli (2020): "Capital Accumulation and Structural Transformation," *The Quarterly Journal of Economics*, 135, 1037–1094.

Christiaensen, L. and W. Martin (2018): "Agriculture, Structural Transformation and Poverty Reduction: Eight New INsights," *World Development*, 109, 413–416.

Colmer, J. (2021): "Temperature, Labor Reallocation, and Industrial Production: Evidence from India," *American economic journal. Applied economics.*

Dar, A. (2019): ""All is Water": Technological Complementarities and Path Dependence in Indian Agriculture," *Working Paper.*

Davis, D. R. and D. E. Weinstein (2002): "Bones, Bombs, and Break Points: the Geography of Economic Activity," *The American Economic Review*, 92, 1269–1289.

Dell, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863–1903.

Dell, M. and B. A. Olken (2020): "The development effects of the extractive colonial economy: The dutch cultivation system in java," *The Review of Economic Studies*, 87, 164–203.

Dell, M. and P. Querubin (2018): "Nation Building through Foreign Intervention: Evidence from Discontinuities in Military Strategies," *The Quarterly Journal of Economics*, 133, 701–764.

Dillon, A. and R. Fishman (2019): "Dams: Effects of Hydrological Infrastructure on Development," *Annual Review of Resource Economics*, 11, 125–148.

Donaldson, D. and A. Storeygard (2016): "The view from above: Applications of satellite data in economics," *The journal of economic perspectives: a journal of the American Economic Association*, 30, 171–198.

Duflo, E. and R. Pande (2007): "Dams," *The Quarterly Journal of Economics*, 122, 601–646.

Eckert, F., M. Peters, et al. (2018): "Spatial Structural Change," *Unpublished Manuscript.*

Egger, P. H. and A. Lassmann (2015): "The Causal Impact of Common Native Language on International Trade: Evidence from a Spatial Regression Discontinuity Design," *The Economic Journal*, 125, 699–745.

Emerick, K. (2018): "Agricultural Productivity and the Sectoral Reallocation of Labor in Rural India," *Journal of Development Economics*, 135, 488–503.

Fajgelbaum, P. and S. Redding (2018): "Trade, structural transformation and development: Evidence from Argentina 1869-1914," *NBER Working Paper*, 20217.

Fishman, R., M. Jain, and A. Kishore (2013): "Groundwater Depletion, Adaptation and Migration: evidence from Gujarat, India," *International Food Policy Research Institute*, 1–39.

Foster, A. D. and M. R. Rosenzweig (2004): "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000," *Economic Development and Cultural Change*, 52, 509–542.

Gollin, D., C. W. Hansen, and A. Wingender (2018): "Two Blades of Grass: The Impact of the Green Revolution," Tech. rep., National Bureau of Economic Research.

Gollin, D., S. Parente, and R. Rogerson (2002): "The Role of Agriculture in Development," *The American Economic Review*, 92, 160–164.

Gollin, D. and R. Rogerson (2014): "Productivity, Transport Costs and Subsistence Agriculture," *Journal of Development Economics*, 107, 38–48.

Gonzalez, R. M. (2021): "Cell Phone Access and Election Fraud: Evidence from a Spatial Regression Discontinuity Design in Afghanistan," *American Economic Journal: Applied Economics*, 13, 1–51.

Government of India (2017): "India on the Move and Churning: New Evidence," .

Hansen, Z. K., G. D. Libecap, and S. E. Lowe (2011): *9. Climate Variability and Water Infrastructure*, University of Chicago Press.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018): "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade," *The Quarterly Journal of Economics*, 133, 357–406.

Henderson, J. V., A. Storeygard, and U. Deichmann (2017): "Has Climate Change driven Urbanization in Africa?" *Journal of development economics*, 124, 60–82.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012): "Measuring Economic Growth from Outer Space," *The American Economic Review*, 102, 994–1028.

Hornbeck, R. and P. Keskin (2014): "The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought," *American Economic Journal: Applied Economics*, 6, 190–219.

——— (2015): "Does Agriculture generate Local Economic Spillovers? Short-run and Long-run Evidence from the Ogallala Aquifer," *American Economic Journal: Economic Policy*, 7, 192–213.

Jain, M., P. Mondal, G. L. Galford, G. Fiske, and R. S. DeFries (2017): "An Automated Approach to Map Winter Cropped Area of Smallholder Farms across Large Scales using MODIS Imagery," *Remote Sensing*, 9, 566.

Johnston, B. F. and J. W. Mellor (1961): "The Role of Agriculture in Economic Development," *The American Economic Review*, 51, 566–593.

Jones, M. R., F. Kondylis, J. A. Loeser, and J. Magruder (2019): *Factor Market Failures and the Adoption of Irrigation in Rwanda*, The World Bank.

Krishnaswamy, N. (2019): "At What Price? Price Supports, Agricultural Productivity, and Misallocation," Tech. rep., Working paper.

Kuznets, S. (1961): "Economic Growth and the Contribution of Agriculture: Notes on Measurement," *International Journal of Agrarian Affairs*, 3.

Lewis, W. A. (1954): "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, 22, 139–191.

Lusome, R. and R. Bhagat (2006): "Trends and patterns of internal migration in India, 1971-2001," in *Annual conference of Indian Association for the Study of Population (IASP) during*, Indian Association for the Study of Population (IASP) Thiruvananthapuram, vol. 7, 9.

Matsuyama, K. (1992): "Agricultural Productivity, Comparative Advantage, and Economic Growth," *Journal of economic theory*, 58, 317–334.

McArthur, J. W. and G. C. McCord (2017): "Fertilizing Growth: Agricultural Inputs and their Effects in Economic Development," *Journal of development economics*, 127, 133–152.

Michaels, G. and F. Rauch (2017): "Resetting the Urban Network: 117–2012," *The Economic Journal*, 128, 378–412.

Michaels, G., F. Rauch, and S. J. Redding (2012): "Urbanization and Structural Transformation," *The Quarterly Journal of Economics*, 127, 535–586.

Michalopoulos, S. and E. Papaioannou (2013): "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81, 113–152.

Nagy, D. K. (2020): "Hinterlands, City Formation and Growth: Evidence from the US Westward Expansion," Tech. rep., Barcelona Graduate School of Economics.

Nunn, N. and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94, 20–36.

Olmstead, S. M. and H. Sigman (2015): "Damming the Commons: An Empirical Analysis of International Cooperation and Conflict in Dam Location," *Journal of the Association of Environmental and Resource Economists*, 2, 497–526.

Pradhan, K. C. (2017): "Unacknowledged Urbanisation: The New Census Towns in India," in *Subaltern urbanisation in India*, Springer, 39–66.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999): "Index that quantifies topographic heterogeneity," *intermountain Journal of sciences*, 5, 23–27.

Rostow, W. W. (1960): *The Stages of Economic Growth*, Cambridge university press.

Ryan, N. and A. Sudarshan (2020): "Rationing the Commons," Tech. rep., National Bureau of Economic Research.

Sekhri, S. (2014): "Wells, Water, and Welfare: the Impact of Access to Groundwater on Rural Poverty and Conflict," *American Economic Journal: Applied Economics*, 6, 76–102.

Smith, C. (2019): "Land Concentration and Long-run Development: Evidence from the Frontier United States," Tech. rep., MIT Working Paper.

Sotelo, S. (2020): "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, 128, 2690–2738.

Strobl, E. and R. O. Strobl (2011): "The Distributional Impact of Large Dams: Evidence from Cropland Productivity in Africa," *Journal of development Economics*, 96, 432–450.

Studwell, J. (2013): *How Asia works: Success and failure in the world's most dynamic region*, Open Road+ Grove/Atlantic.

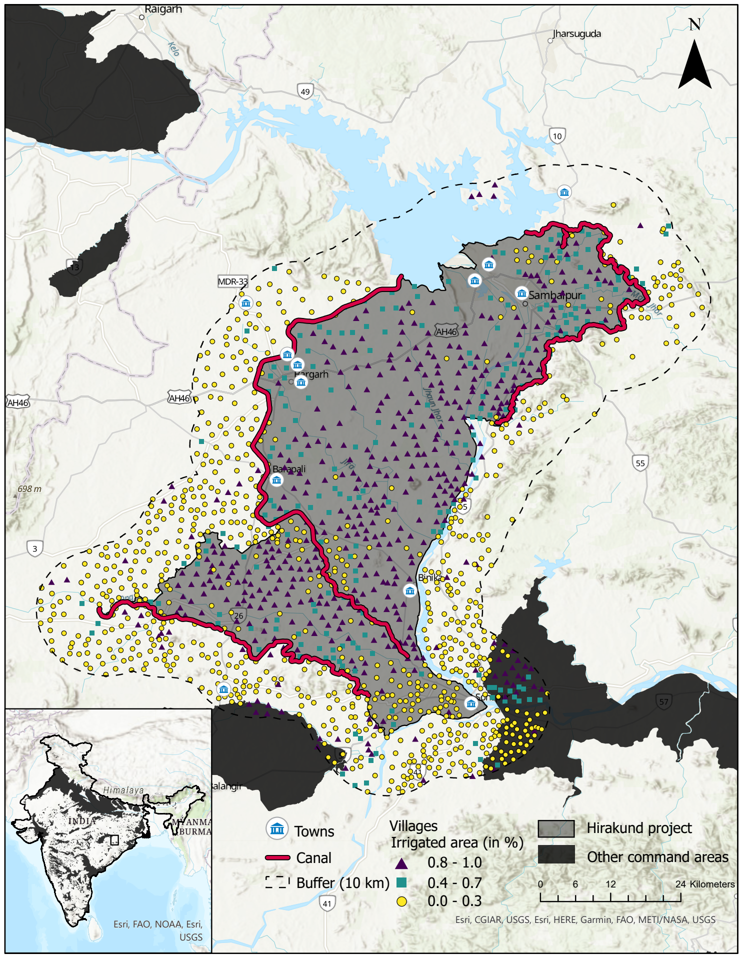
Sukhtankar, S. (2016): "Does Firm Ownership Structure Matter? Evidence from Sugar Mills in India," *Journal of Development Economics*, 122, 46–62.

von der Goltz, J., A. Dar, R. Fishman, N. D. Mueller, P. Barnwal, and G. C. McCord (2020): "Health Impacts of the Green Revolution: Evidence from 600,000 births across the Developing World," *Journal of health economics*, 74, 102373.

Zaveri, E., J. Russ, and R. Damania (2020): "Rainfall Anomalies are a Significant Driver of Cropland Expansion," *Proceedings of the National Academy of Sciences*, 117, 10225–10233.

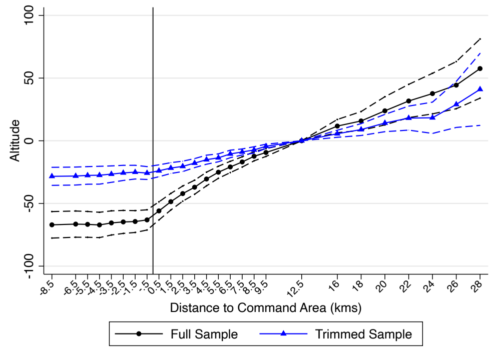
Figures

Figure 1: Illustration of a Canal Command Area (Hirakud Major Irrigation Project)

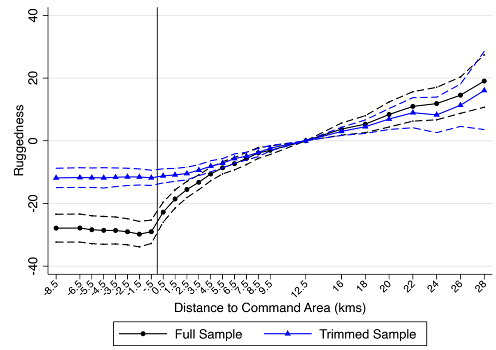


Notes: The empirical strategy compares villages on either side of the command area border (shaded light grey) in a 10 km buffer (denoted by the dotted black line). To compare nearby villages, 5 km boundary segment fixed effects are used, which are calculated by splitting the border into smaller parts. (Boundary segments not shown for simplicity.) The estimating sample is restricted to parts of the border which have a slope less than 1.5 degrees on the outside of the border. (This sample restriction gives us a balanced sample on key geographic variables. See Figure 2.) This map illustrates the two types of estimation samples that are used in the study: the main results use the entire canal command area boundary, with the caveats mentioned above. A second estimation sample, used in robustness checks, relies only on the part of the command area boundary that is contiguous with the canal. In this example, only villages on either side of the command area border (black solid line) which overlaps with the canal (red solid line) will be used.

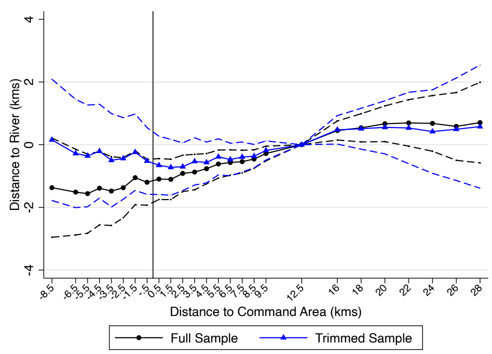
Figure 2: Geographic Features



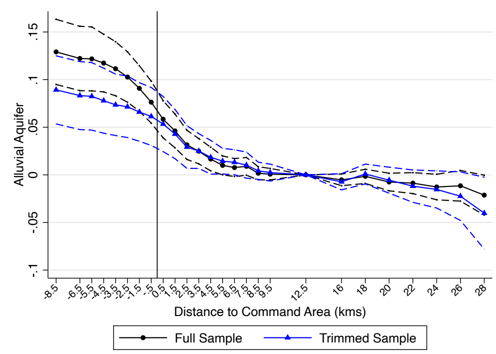
2.1: Altitude



2.2: Ruggedness



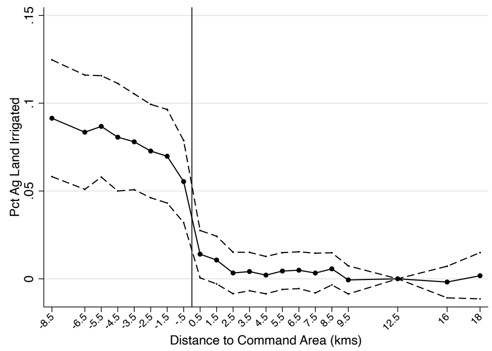
2.3: Distance to River



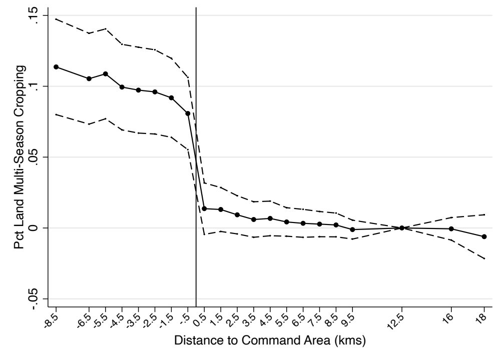
2.4: Aquifer

Notes: This figure compares key geographic features in villages inside the command area (to the left of 0) with those just outside (to the right of 0). Distance to the command area (in km) is on the x-axis. The solid line represents results from a regression of pre-determined, geographic characteristics on canal command area treatment dummy, binned distances, controls and 5 km boundary segment fixed effects. Standard errors are clustered at the project code level. The dotted lines illustrate the 95 percent confidence intervals. The black lines refers to the full sample while the blue lines refers to the restricted/trimmed sample (see definitions in text). Figure 2.1 depicts altitude (in meters), Figure 2.2 depicts the terrain ruggedness index derived from USGS digital elevation models, Figure 2.3 depicts distance to river (in kms), and Figure 2.4 depicts whether a village lies on top of an alluvium/water-deposited aquifer.

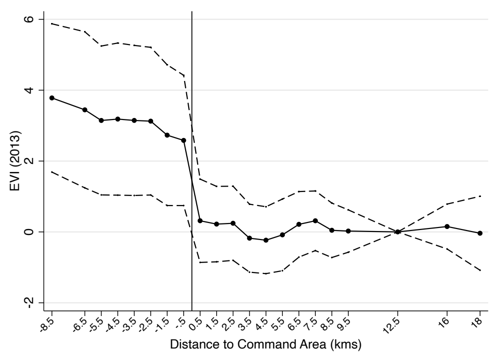
Figure 3: Agriculture and Development



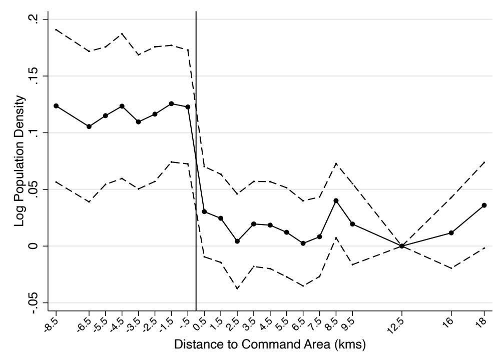
3.1: Pct of Agriculture Area Irrigated



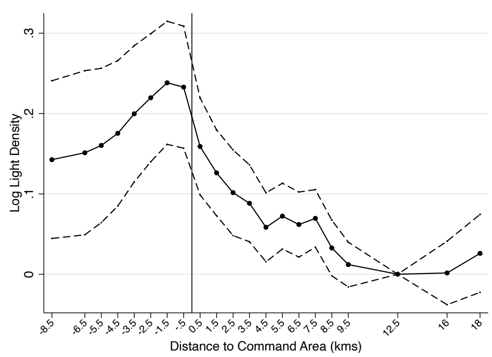
3.2: Multi-Season Cropping



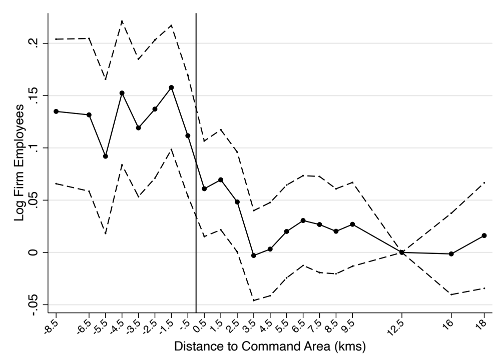
3.3: Dry Season Vegetation



3.4: Log Population Density



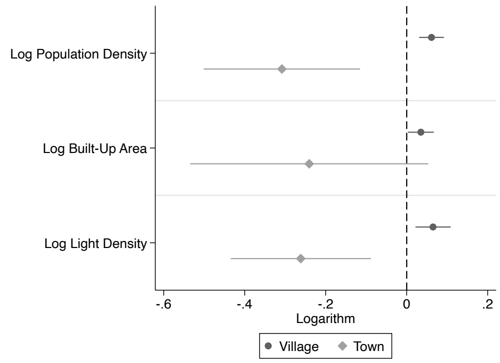
3.5: Log Light Density



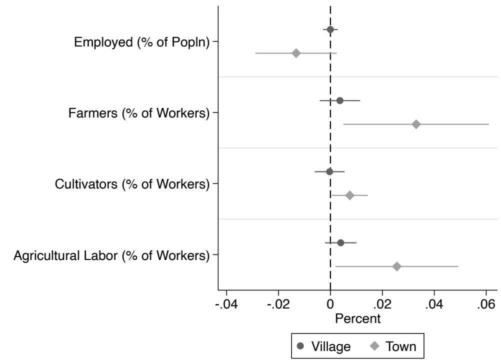
3.6: Log Firm Employment

Notes: This figure compares agricultural and development outcomes in villages inside the command area (to the left of 0) with those just outside (to the right of 0). Distance to the command area (in km) is on the x-axis. The solid line represents results from a regression of outcomes on canal command area treatment dummy, binned distances, controls and 5 km boundary segment fixed effects. Standard errors are clustered at the project code level. The dotted lines represent 95% confidence intervals. Figure 3.1 depicts area under irrigation as percent of cultivable land; Figure 3.2 depicts land area that is cropped twice or thrice as percentage of agricultural area; and Figure 3.3 depicts dry season vegetation indices as percentage of total village area. Figure 3.5 depicts mean nighttime lights per sq km. Figure 3.6 depicts number of employees in firms across manufacturing, agriculture and services enterprises.

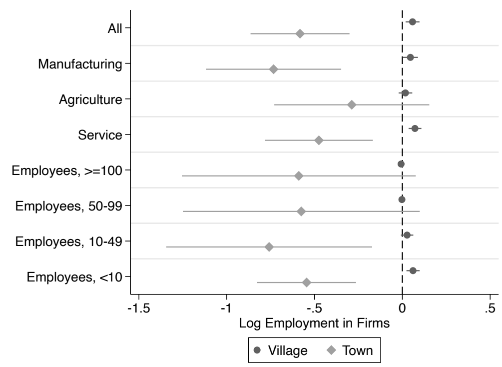
Figure 4: Labor Force Participation, Firm Activity and Assets



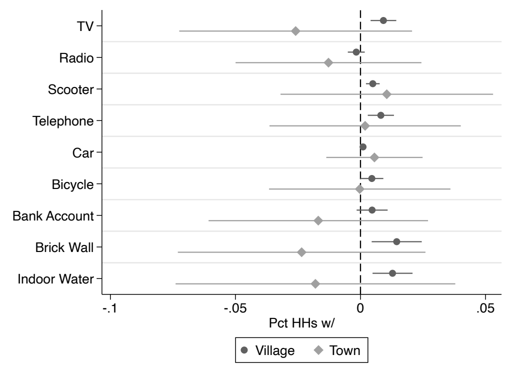
4.1: Urbanization



4.2: Labor Force Participation



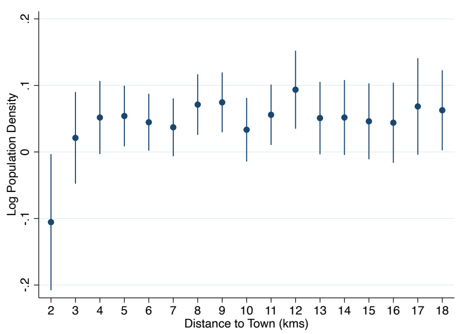
4.3: Employment in Firms



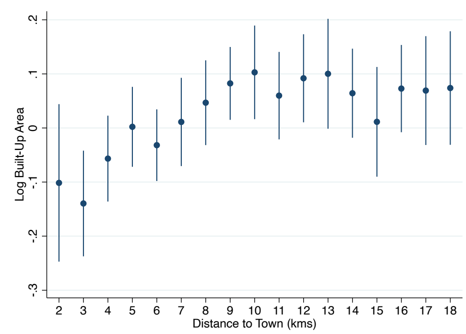
4.4: Assets

Notes: This figure plots β from equation 1 for non-agricultural outcomes in villages and towns. Figure 4.2 depicts the impact on labor force participation (Census of India 2011): employed refers to workers as % of population; farmers refers to sum of cultivators and agricultural laborers as a % of all workers; cultivators refers to those directly involved in farming or supervision of farming, and unlike agricultural labors they work on their own farm. Figure 4.3 depicts ln(employment) in firms by sector and firm size (Economic Census 2012-13). All refers to sum of workers employed in manufacturing, agriculture and services enterprises. Sectors are classified using Ministry of Statistics and Programme Implementation's National Industrial Classification. Firm size is measured using number of workers: employees, $\geq$ 100, 50-99, 10-49 and $<$ 10 refers to firms with more than 100 workers, between 50 and 99 workers, between 10-49 workers and less than 10 workers respectively. Figure 4.4 depicts assets and amenities as % of households in villages/towns (Census of India 2011).

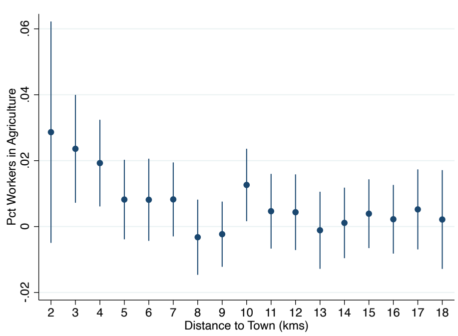
Figure 5: Treatment Effect by Distance to Town



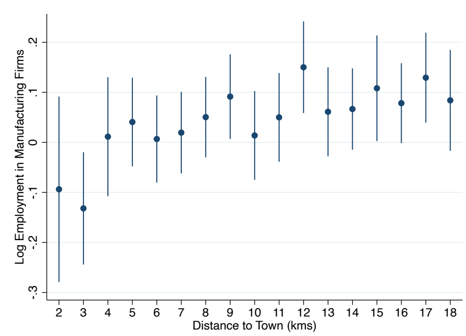
5.1: Log Population Density



5.2: Log Built-Up Area



5.3: Pct Farmers



5.4: Log Employment in Manu. Firms

Notes: The above figure plots the coefficient on the interaction of the treatment dummy in villages with distance to towns for urbanization, agricultural and non-agricultural outcomes. Figure 5.1, Figure 5.2, 5.3 and Figure 5.4 depict heterogeneous effects for population, built-up area, farmers and employment in manufacturing firms. Definitions same as before.

Tables

Table 1: Agriculture (Census)

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Pct Ag Area Irrigated Canal (Census 2011)				
Treatment	0.107***	0.084***		
	(0.009)	(0.008)	NA	
Control Mean	0.051			
R-squared	0.249	0.376		
N	145475	142951		
Panel B: Pct Ag Area Irrigated (Census 2011)				
Treatment	0.070***	0.056***		
	(0.008)	(0.007)	NA	
Control Mean	0.417			
R-squared	0.576	0.680		
N	145581	143059		
Project FE	Yes			
Boundary Segment FE		Yes		
District FE	Yes	Yes		

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (column 1) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (column 2) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. Agricultural outcomes are derived from Census of India 2011. Data is available only for villages and not for towns. Panel A reports area irrigated using canals (as percentage of cultivable area); and panel B reports total area irrigated by all sources, surface- or ground-water (as percentage of cultivable area). Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * p<0.10, ** p<0.05, *** p<0.01.

Table 2: Agriculture (Remotely-sensed)

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Pct Area Cultivated (2011-12)				
Treatment	0.080***	0.070***	0.120***	0.168***
	(0.006)	(0.005)	(0.039)	(0.049)
Control Mean	0.591		0.333	
R-squared	0.506	0.649	0.633	0.728
N	145609	143087	1513	791
Panel B: Pct Area Multi-Season Cropping (2011-12)				
Treatment	0.089***	0.073***	0.093***	0.117**
	(0.008)	(0.007)	(0.031)	(0.046)
Control Mean	0.286		0.168	
R-squared	0.571	0.720	0.601	0.710
N	144240	141742	1479	775
Panel C: EVI (2013)				
Treatment	2.792***	2.839***	1.132	1.889*
	(0.530)	(0.543)	(0.842)	(1.017)
Control Mean	15.896		7.293	
R-squared	0.734	0.830	0.764	0.814
N	125028	122485	1439	748
Project FE	Yes		Yes	
Boundary Segment FE		Yes		Yes
District FE	Yes	Yes	Yes	Yes

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1 and 3) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2 and 4) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. Agricultural outcomes are derived from satellite data: panel A reports area cultivated from NRSC/ISRO 2011-12; panel B reports area cropped twice or thrice in a year, also from NRSC/ISRO 2011-12; and panel C reports dry-season vegetation from MODIS EVI 2013. All remotely sensed data are measured as percentage of total area. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3: Urbanization

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Log Population Density				
Treatment	0.070***	0.061***	-0.200**	-0.308***
	(0.014)	(0.016)	(0.080)	(0.098)
Control Mean	5.715		7.766	
R-squared	0.421	0.488	0.513	0.606
N	136879	134305	1467	781
Panel B: Log Light Density				
Treatment	0.086***	0.065***	-0.137	-0.261***
	(0.024)	(0.022)	(0.088)	(0.088)
Control Mean	1.378		3.117	
R-squared	0.535	0.743	0.605	0.831
N	133030	130487	1440	759
Panel C: Log Built Up Area				
Treatment	0.032**	0.035**	-0.153*	-0.268*
	(0.014)	(0.016)	(0.086)	(0.152)
Control Mean	6.777		9.304	
R-squared	0.299	0.387	0.663	0.765
N	109185	106386	1411	759
Project FE	Yes		Yes	
Boundary Segment FE		Yes		Yes
District FE	Yes	Yes	Yes	Yes

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1 and 3) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2 and 4) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are derived from census and satellite data: panel A reports ln(population density) from Census of India 2011; panel B reports ln(mean nighttime luminosity score per sq km) from NOAA 2013; and panel C reports ln(built up area) from NRSC/ISRO 2011-12. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Workers

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Pct Popln Employed				
Treatment	0.000	0.000	0.003	-0.013*
	(0.001)	(0.001)	(0.005)	(0.008)
Control Mean	0.447		0.421	
R-squared	0.447	0.525	0.606	0.696
N	136879	134305	1387	757
Panel B: Pct Workers Farmers				
Treatment	0.007	0.004	0.032**	0.033**
	(0.004)	(0.004)	(0.013)	(0.014)
Control Mean	0.767		0.135	
R-squared	0.324	0.463	0.601	0.716
N	136883	134309	1387	757
Panel C: Pct Workers Own-Farm				
Treatment	-0.002	0.000	0.010***	0.007**
	(0.003)	(0.003)	(0.003)	(0.004)
Control Mean	0.349		0.040	
R-squared	0.332	0.430	0.576	0.634
N	136883	134309	1387	757
Panel D: Pct Workers Ag Labor				
Treatment	0.009***	0.004	0.022**	0.026**
	(0.003)	(0.003)	(0.010)	(0.012)
Control Mean	0.418		0.096	
R-squared	0.340	0.433	0.601	0.722
N	136883	134309	1387	757
Project FE	Yes		Yes	
Boundary Segment FE		Yes		Yes
District FE	Yes	Yes	Yes	Yes

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1 and 3) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2 and 4) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are derived from Census of India 2011: panel A reports total employment (as percent of population); panel B reports farmers (as percent of workers); panel C reports own-farm workers/cultivators (as percent of workers); and panel D reports agricultural laborers (as percent of workers). Farmers = own-farm workers/cultivators + ag laborers. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5: Firms

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Log Employees				
Treatment	0.066***	0.058***	-0.263**	-0.583***
	(0.021)	(0.020)	(0.133)	(0.142)
Control Mean	3.760		7.577	
R-squared	0.465	0.544	0.506	0.626
N	128402	125796	1467	781
Panel B: Log Manu Employees				
Treatment	0.060**	0.046**	-0.322*	-0.733***
	(0.026)	(0.022)	(0.172)	(0.195)
Control Mean	1.664		6.045	
R-squared	0.310	0.418	0.516	0.653
N	128402	125796	1467	781
Panel C: Log Ag Employees				
Treatment	0.028	0.018	-0.086	-0.288
	(0.024)	(0.020)	(0.133)	(0.223)
Control Mean	1.635		3.671	
R-squared	0.594	0.675	0.623	0.727
N	128402	125796	1467	781
Panel D: Log Service Employees				
Treatment	0.074***	0.072***	-0.231**	-0.475***
	(0.017)	(0.019)	(0.110)	(0.155)
Control Mean	3.170		7.029	
R-squared	0.359	0.446	0.517	0.610
N	128402	125796	1467	781
Panel E: Log Employees >100 Workers				
Treatment	-0.010	-0.007	-0.604***	-0.590*
	(0.007)	(0.007)	(0.230)	(0.337)
Control Mean	0.081		1.850	
R-squared	0.067	0.200	0.366	0.502
N	128402	125796	1467	781
Panel F: Log Employees 50-99 Workers				
Treatment	-0.002	-0.002	-0.563***	-0.576*
	(0.005)	(0.006)	(0.162)	(0.341)
Control Mean	0.096		2.139	
R-squared	0.134	0.254	0.410	0.529
N	128402	125796	1467	781
Panel G: Log Employees 10-49 Workers				
Treatment	0.035***	0.027	-0.328	-0.758**
	(0.013)	(0.018)	(0.210)	(0.296)
Control Mean	0.661		4.700	
R-squared	0.232	0.331	0.464	0.546
N	128402	125796	1467	781
Panel H: Log Employees <10 Workers				
Treatment	0.065***	0.061***	-0.208*	-0.545***
	(0.022)	(0.019)	(0.125)	(0.142)
Control Mean	3.673		7.350	

Continued on next page

Table 5 – Continued from previous page

R-squared	0.471	0.550	0.525	0.636
N	128402	125796	1467	781
Project FE	Yes		Yes	
Boundary Segment FE		Yes		Yes
District FE	Yes	Yes	Yes	Yes

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1 and 3) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2 and 4) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are derived from Economic Census 2012-13: panel A reports ln(total employment) in all enterprises/firms. Total employment = agriculture + manufacturing + services. Panel B reports ln(manufacturing sector employment); panel C reports ln(agricultural sector employment); panel D reports ln(service sector employment). While panel B to panel D report sectoral impacts, panel E to panel H report impacts by firm size: panel E, F, G and H report ln(employment) for firms with greater than 100 workers, between 50-99 workers, 10-49 workers and less than 10 workers respectively. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: Assets and Housing

	Villages		Towns	
	(1)	(2)	(3)	(4)
Panel A: Pct w/TV				
Treatment	0.010***	0.009***	-0.014	-0.026
	(0.003)	(0.003)	(0.013)	(0.024)
Control Mean	0.268		0.632	
R-squared	0.697	0.758	0.745	0.840
N	136273	133720	1467	781
Panel B: Pct w/Radio				
Treatment	-0.003	-0.002	-0.005	-0.013
	(0.002)	(0.002)	(0.007)	(0.019)
Control Mean	0.159		0.209	
R-squared	0.266	0.337	0.712	0.748
N	136273	133720	1467	781
Panel C: Pct w/Scooter				
Treatment	0.006***	0.005***	-0.000	0.011
	(0.001)	(0.001)	(0.012)	(0.021)
Control Mean	0.137		0.262	
R-squared	0.550	0.625	0.698	0.832
N	136273	133720	1467	781
Panel D: Pct w/Telephone				
Treatment	0.009***	0.008***	-0.006	0.002
	(0.003)	(0.003)	(0.010)	(0.019)
Control Mean	0.504		0.712	
R-squared	0.476	0.545	0.674	0.798
N	136273	133720	1467	781
Panel E: Pct w/Car				
Treatment	0.001***	0.001**	0.003	0.006
	(0.000)	(0.000)	(0.004)	(0.010)
Control Mean	0.016		0.047	
R-squared	0.215	0.291	0.552	0.711
N	136273	133720	1467	781
Panel F: Pct w/Bicycle				
Treatment	0.009***	0.005*	0.003	-0.000
	(0.003)	(0.002)	(0.011)	(0.018)
Control Mean	0.495		0.509	
R-squared	0.591	0.663	0.707	0.825
N	136273	133720	1467	781
Panel G: Pct w/Banking				
Treatment	0.005	0.005	-0.008	-0.017
	(0.003)	(0.003)	(0.011)	(0.022)
Control Mean	0.529		0.596	
R-squared	0.375	0.472	0.536	0.654
N	136273	133720	1467	781
Panel H: Pct w/Brick Wall				
Treatment	0.014***	0.014***	-0.019	-0.024
	(0.005)	(0.005)	(0.013)	(0.025)
Control Mean	0.446		0.737	
R-squared	0.608	0.691	0.709	0.736
N	136273	133720	1467	781

Continued on next page

Table 6 – Continued from previous page

Panel I: Pct w/Inside Water				
Treatment	0.020***	0.013***	0.006	-0.018
	(0.004)	(0.004)	(0.014)	(0.028)
Control Mean	0.281		0.539	
R-squared	0.541	0.629	0.743	0.825
N	136273	133720	1467	781
Panel J: Pct w/Condition Good				
Treatment	0.011***	0.010***		
	(0.003)	(0.003)		
Control Mean	0.427		NA	
R-squared	0.222	0.305		
N	136273	133720		
Panel K: Number Rooms				
Treatment	0.040***	0.036***		
	(0.007)	(0.007)		
Control Mean	2.874		NA	
R-squared	0.516	0.592		
N	136273	133720		
Project FE	Yes		Yes	
Boundary Segment FE		Yes		Yes
District FE	Yes	Yes		Yes

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1 and 3) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2 and 4) where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; η_p are project fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are derived from Census of India 2011 and are reported as percentage of households. Definitions for outcomes in panel A to panel I, and panel K are self explanatory. Panel J reports percentage of households who report that their house is in a 'good' condition (as opposed to 'livable' or 'dilapidated').

Table 7: Urbanization in Villages (by Proximity to Town)

	Log		
	Population Density	Built-up Area	Light Density
	(1)	(2)	(3)
Treatment	0.067***	0.056***	0.071***
	(0.015)	(0.016)	(0.021)
Prox Town	0.159***	0.355***	0.623***
	(0.017)	(0.027)	(0.029)
Treat × Prox Town	-0.050**	-0.147***	-0.068
	(0.023)	(0.034)	(0.044)
R-squared	0.489	0.390	0.760
N	134305	106386	130487

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \delta Prox\ Town + \kappa (C_i \times Prox\ Town_i) + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; $Prox\ Town_i$ is a binary variable taking a value of 1 if village i is within 4 kms distance to a town, $C_i \times Prox\ Town_i$ is the interaction of the two indicator variables; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are derived from census and satellite data: column (1) reports ln(population density) from Census of India 2011; column (2) reports ln(built up area) from NRSC/ISRO 2011-12; and column (3) reports ln(mean nighttime luminosity score per sq. km) from NOAA 2013. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 8: Labor Force in Villages (by Proximity to Town)

	Pct		Log		
	Population Workers	Farmers	All Workers	Farmers	Non-Ag Workers
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.001	0.002	0.065***	0.073***	0.068***
	(0.001)	(0.003)	(0.015)	(0.014)	(0.025)
Prox Town	-0.017***	-0.093***	0.121***	-0.080***	0.459***
	(0.002)	(0.006)	(0.018)	(0.020)	(0.029)
Treat X Prox Town	0.007***	0.021***	-0.038	0.034	-0.085**
	(0.002)	(0.007)	(0.024)	(0.026)	(0.035)
R-squared	0.525	0.471	0.532	0.555	0.476
N	134305	134309	134309	133936	131189

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \delta Prox\ Town + \kappa(C_i \times Prox\ Town_i) + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; $Prox\ Town_i$ is a binary variable taking a value of 1 if village i is within 4 kms distance to a town, $C_i \times Prox\ Town_i$ is the interaction of the two indicator variables; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes come from the Census of India 2011. Column (1) reports workers who are employed (as percent of population); column (2) reports farmers (as percent of total workers). Farmers = cultivators + agricultural laborers. Column (3) reports ln(total number of workers); column (4) refers to ln(farmers); and column (5) reports ln(non-agricultural workers). All workers = farmers + non-agricultural workers. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * p<0.10, ** p<0.05, *** p<0.01.

Table 9: Firms in Villages (by Proximity to Town)

	Log Employment							
	Sector				Size			
					Number Workers			
	All	Manu	Ag	Service	> 100	50-99	10-49	<10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.065***	0.058***	0.015	0.081***	-0.007	-0.001	0.029*	0.068***
	(0.019)	(0.020)	(0.020)	(0.018)	(0.006)	(0.006)	(0.017)	(0.019)
Prox Town	0.227***	0.265***	0.019	0.243***	0.053***	0.070***	0.194***	0.192***
	(0.026)	(0.033)	(0.025)	(0.025)	(0.016)	(0.017)	(0.030)	(0.024)
Treat X Prox Town	-0.059*	-0.095**	0.021	-0.073**	-0.001	-0.014	-0.020	-0.058*
	(0.035)	(0.043)	(0.031)	(0.032)	(0.021)	(0.022)	(0.037)	(0.032)
R-squared	0.545	0.420	0.675	0.447	0.200	0.254	0.332	0.550
N	125796	125796	125796	125796	125796	125796	125796	125796

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \delta Prox\ Town + \kappa(C_i \times Prox\ Town_i) + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; $Prox\ Town_i$ is a binary variable taking a value of 1 if village i is within 4 kms distance to a town, $C_i \times Prox\ Town_i$ is the interaction of the two indicator variables; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The outcomes are from Economic Census 2012-13. Columns (1)-(4) report impacts by sector. Column (1) reports ln(employment) across all enterprises/firms. All refers to sum of workers employed in manufacturing, agriculture and services enterprises. Column (2), (3) and (4) report ln(employment) in manufacturing, agriculture and service sector respectively. Sectors are classified using Ministry of Statistics and Programme Implementation's National Industrial Classification. Columns (5)-(8) report impacts by firm size: greater than 100 workers (column 5), between 50-99 workers (column 6), between 10-49 workers (column 7) and less than 10 workers (column 8). Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * p<0.10, ** p<0.05, *** p<0.01.

Table 10: Assets and Housing, Villages (by town proximity)

	Assets							Housing			
	Pct						Pct Bank	Pct			Num
	TV	Radio	Scooter	Phone	Car	Bicycle	Account	Brick Wall	Inside Water	Condition Good	Rooms
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment	0.011***	-0.002	0.006***	0.011***	0.001***	0.006**	0.006**	0.017***	0.014***	0.010***	0.041***
	(0.003)	(0.002)	(0.001)	(0.003)	(0.000)	(0.002)	(0.003)	(0.005)	(0.004)	(0.003)	(0.007)
Prox Town	0.047***	-0.001	0.017***	0.030***	0.003***	0.024***	0.006	0.050***	0.036***	0.022***	0.046***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.001)	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.009)
Treat X Prox Town	-0.015***	-0.000	-0.010***	-0.019***	-0.001	-0.008*	-0.014**	-0.018***	-0.008*	-0.002	-0.043***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.001)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.012)
R-squared	0.759	0.337	0.625	0.545	0.291	0.663	0.472	0.692	0.630	0.305	0.592
N	133720	133720	133720	133720	133720	133720	133720	133720	133720	133720	133720

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \delta Prox\ Town + \kappa(C_i \times Prox\ Town_i) + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; $Prox\ Town_i$ is a binary variable taking a value of 1 if village i is within 4 kms distance to a town, $C_i \times Prox\ Town_i$ is the interaction of the two indicator variables; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. The asset and amenities outcomes are derived from Census of India and calculated as % of households in villages. Variables are self-explanatory. House in good condition which refers to whether a house is reported as 'good' (as opposed to 'livable' or 'dilapidated'). Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * p<0.10, ** p<0.05, *** p<0.01.

A1 Appendix: Theoretical Framework

This section characterizes the equilibrium of the model presented in the main body in detail. We start by solving the problem of landowners. We then show the evolution of prices in the rest of the world. Lastly, we present the equations that define the equilibrium in the domestic economy. To save on notation, we drop index of region and time in what follows, unless otherwise indicated.

Landowners. In each region, total land consists of a continuum of plots $\ell \in L$, where L is the total area of the region. There is a landowner in each region who chooses how to assign plots between agriculture and manufacturing. To be assigned to sector k, landowners have to incur a conversion cost of $e_k(\ell)$, proportional to land rents. The maximization problem of the landowner is thus

$$\max_k r_k e_k(\ell).$$

We assume that conversion costs are drawn from a Fréchet, $F(\epsilon) = 1 - exp(-\epsilon^{-\theta})$. We can derive the share of land employed in activity k as follows

$$\begin{aligned}
\lambda_k &= \int_0^\infty P\left(r_k e_k > r_{k-} e_{k-}\right) \exp\left(-e_k^{-\theta}\right) \theta e_k^{-\theta-1} de_k \\
&= \int_0^\infty P\left(e_k \frac{r_k}{r_{k-}} > e_{k-}\right) \exp\left(-e_k^{-\theta}\right) \theta e_k^{-\theta-1} de_k \\
&= \int_0^\infty \exp\left(-\left(e_k \frac{r_k}{r_{k-}}\right)^{-\theta}\right) \exp\left(-e_k^{-\theta}\right) \theta e_k^{-\theta-1} de_k \\
&= \int_0^\infty \exp\left(-e_k^{-\theta}\left(\frac{r_{k-}^\theta + r_k^\theta}{r_k^\theta}\right)\right) \theta e_k^{-\theta-1} de_k \\
&= \left(\frac{r_k^\theta}{r_{k-}^\theta + r_k^\theta}\right) \int_0^\infty \exp\left(-\left(\frac{r_{k-}^\theta + r_k^\theta}{r_k^\theta}\right) e_k^{-\theta}\right) \theta \left(\frac{r_{k-}^\theta + r_k^\theta}{r_k^\theta}\right) e_k^{-\theta-1} de_k \\
&= \left(\frac{r_k^\theta}{r_{k-}^\theta + r_k^\theta}\right)
\end{aligned}$$

Market Equilibrium in the Foreign Economy. Since workers are freely mobile, marginal productivity must equalize between sectors

$$p_A A_A \left(1-\alpha_A\right) \left(L_A\right)^{\alpha_A} \left(N_A\right)^{-\alpha_A} = A_M \left(1-\alpha_M\right) \left(L_M\right)^{\alpha_M} \left(N_M\right)^{-\alpha_M}.$$

Rearranging the equation gives

$$p_A = \frac{A_M}{A_A} \frac{\left(1-\alpha_M\right)}{\left(1-\alpha_A\right)} \frac{\left(L_M/N_M\right)^{\alpha_M}}{\left(L_A/N_A\right)^{\alpha_A}}.$$

Let us now obtain L_M, L_A, N_A, and N_M as a function of the parameters. First, labor market clearing gives

$$N_k = N\mu_k.$$

For land employment, first use the following expressions coming from the FOC of firms

$$\frac{N_k}{L_k} = \frac{1-\alpha_k}{\alpha_k}\frac{r_k}{w}.$$

The maximization problem of landowners gives

$$\frac{r_A}{r_M} = \left(\frac{L_A}{L_M}\right)^{\frac{1}{\theta}}. \tag{A1}$$

Combining the three expressions above, after some tedious algebra, we get

$$\frac{L_M}{L_A} = \left(\frac{\mu_M}{\mu_A}\frac{1-\alpha_A}{\alpha_A}\frac{\alpha_M}{1-\alpha_M}\right)^{\frac{\theta}{1+\theta}} \tag{A2}$$

which together with $L = L_A + L_M$ characterizes the optimal allocation of land as an explicit function of parameters. Optimal labor and land allocation are therefore constant over time. Let L_A^*, L_M^*, N_A^*, N_M^* be the equilibrium values in Foreign. The price of agricultural produce at time

t in the rest of the world, indexed by F, is

$$p_{FAt} = \frac{A_{FMt}}{A_{FAt}} \frac{(1-\alpha_M)}{(1-\alpha_A)} \frac{(L_M^*/N_M^*)^{\alpha_M}}{(L_A^*/N_A^*)^{\alpha_A}}.$$

Let n_{FM}^* be the optimal share of workers in manufacturing, the evolution of A_{FMt} is then $A_{FMt+1} - A_{FMt} = (n_{FM}^*)^\gamma$.

Market Equilibrium in the Domestic Economy. We now turn to the domestic economy. Sectoral prices in Town are the same as in the ROW. In Village, to define sectoral prices, we need to define its trade patterns. Given autarky price $p_{VA,t}^A$, Village is an exporter of agricultural goods if $p_{VA,t}^A \tau < p_{FA,t}$, in which case we have $p_{VA,t} = \tau p_{FA,t}$ and $p_{VM,t} = \frac{1}{\tau}$, an importer of agricultural goods if $p_{FA,t}\tau < p_{VA,t}^A$, in which case we have $p_{FA,t} = \frac{1}{\tau} p_{VA,t}$, and in autarky if $1/\tau < p_{VA,t}^A / p_{FA,t} < \tau$, in which case we have $p_{VA,t} = p_{VA,T}^A$.

Given sectoral prices, we now define the equations that characterize the equilibrium in terms of the price of factors of production (rents and wages) and the of workers and land between sectors and regions. First, marginal productivity of labor and land gives

$$r_{ik} = p_{ik} A_{ik} \alpha_k \left(\frac{N_{ik}}{L_{ik}} \right)^{1-\alpha_k} \tag{A3}$$

$$w_i = p_{ik} A_{ik} \left(1-\alpha_k\right) \left(\frac{L_{ik}}{N_{ik}} \right)^{\alpha_k} \tag{A4}$$

Second, first order conditions of firms give

$$N_{ik} = \frac{(1-\alpha_k)}{\alpha_k} \frac{r_{ik} L_{ik}}{w_i}. \tag{A5}$$

Third, the optimal allocation of landowners gives

$$L_{ik} = \frac{r_{ik}^\theta}{r_{iA}^\theta + r_{iM}^\theta} L. \tag{A6}$$

Fourth, workers are fully employed

$$N = \sum_i \sum_k N_{ik}. \tag{A7}$$

Fifth, workers are indifferent between Village and Town

$$\frac{w_V}{p_V} = \frac{w_T}{p_T}, \tag{A8}$$

where $p_i = p_{iA}^{\mu_A} p_{iM}^{\mu_M}$ is the consumer price index. Using equations (A3) to (A8), we can solve for the endogenous variables of the model.

Numerical Example. Figure A1 shows a numerical example of the model developed above for parameter values $\{\theta, \gamma, \alpha_A, \alpha_M, \mu_k, \tau\}$= $\{2, 0.1, 0.5, 0.2, 0.5, 1.05\}$, agricultural productivities $A_{FA}, = A_{TA}, = A_{VA} = 1$, initial conditions for manufacturing productivities $A_{FM0} = A_{TM0} = 1$ and $A_{VM0} = 0.8$, land endowment $L_V = L_T = 0.5$, population $N_H = N_F = 1$. These parameters ensure full specialization of Village in any period of time. We illustrate the impact of an agricultural productivity shock by increasing A_{iA} by 10%.

Appendix Figure A1.1 shows that the agricultural productivity shock has a permanent negative effect on the population in Town. With the agricultural productivity shock, the manufacturing sector shrinks, and the region gains comparative advantage in agriculture, which has no productivity growth over time. That puts Town in a path of permanent lower economic growth and smaller incentives for the inflow of workers. In Village, on the other hand, the opposite happens. The agricultural productivity shock holds the outflow of workers and puts the economy on a path with larger population. This happens, in part, because of the general equilibrium effects in Town: with more workers in Village, there are fewer workers in Town, the price of land in Town drops, which makes Town specialize in agriculture and enter in a path of lower economic growth.

Figure A1.2 and Figure A1.3 show the effect on agricultural employment and land. In Town, the productivity shock generates a permanent effect on the share of workers and land in agriculture. This gap widens over time. In Village, on the other hand, the productivity shock generates an initial expansion, but the effect of the initial shock is attenuated over time since the region becomes fully specialized in agriculture both in the couterfactual and the factual scenario. In the case of Village, the productivity shock basically anticipates the specialization of the region in agriculture. These numerical examples suggest that the impact on the share of employment in agriculture should be smaller in Village (relative to Town), which is indeed something that we observe in the data.

As expected, the impact of the agricultural productivity depends on initial conditions. If, for example, $A_{FM0} = A_{TM0} = A_{VM0} = 1$ and $A_{FA}, = A_{TA}, = A_{VA} = 1$, then all regions have the same comparative advantage. Since there are no incentives for trade in this case, all regions have the same relative price of agricultural goods and keep a constant share of workers in manufacturing and total population in the absence of the shock. Here, an agricultural productivity shock has a negative effect on total population both in Village and in Town. Interestingly, if $A_{FM0} = A_{TM0} = A_{VM0} = 1$ and $A_{FA} = A_{TA} = 1$ and $A_{VA} = 0.9$, then Town and Village have a larger share of workers in manufacturing in the initial period relative to the rest of the world ($n_{VM0} > n^*_{FM}$ and $n_{TM0} > n^*_{FM}$). Relative to the ROW, Town specializes in manufacturing because of its larger population density and Village specializes in manufacturing because of its relative productivities. In that case, a positive productivity shock in agriculture generates a negative effect on the evolution of total population both in Town and in Village.

A2 Appendix: Data and Background

Census of India. The Census of India is a population-wide enumeration exercise conducted in the country every ten years. It publishes data on demographics, economic activity, educational attainment, migration, fertility and household amenities and assets for the entire country. We use three 'series' of the census in this paper that are aggregated at the village and town level: (i) A-Series: General Population; (ii) B-Series: Economic Tables; (iii) H-Series: Houses, Household Amenities and Assets Tables.

From the A-Series, we extract data on total population in a village/town, population of Scheduled Castes (SCs) and population of Scheduled Tribes (STs).[27] From the B-Series, we use data to classify workers as those engaged in agricultural or non-agricultural practices. The census distinguishes between workers according to: (a) whether workers worked more than half of the months in a year viz. 'main' ($\geq$ 6 months) and 'marginal' ($<$ 6 months) workers; (b) type of work which is categorized in 4 ways viz cultivators, agricultural laborers, household industry workers and others; and (c) sector of employment which is categorized in 9 ways viz. agricultural and allied activities, mining and quarrying, manufacturing, electricity, gas and water supply, construction, wholesale, retail trade and repair work, hotel and restaurants, transport, storage and communications, financial intermediation, real estate, business activities, and other services.

In 2011, there were 481.7 million workers in the country, out of which 118.7 million were cultivators, 144.3 million agricultural laborers, 18.3 million household industry workers and 200.4 million other types of workers. Cultivators are defined as those who are directly engaged in farming or involved in the supervision of farm activities.[28] Agricultural laborers are those who worked someone else's land in exchange for wages either in cash or kind. Household industry workers refer to those who are involved in the production, processing, servicing, repairing or making and selling of goods, as long as the 'industry' involved members of household and run on a small scale and not that of a factory.

[27] SCs and STs are the most marginalized communities in the country.

[28] Farming is defined as ploughing, sowing and harvesting cereals, millets, pulses or fibre crops. The cultivation of fruits, vegetables, growing orchards/groves or working on plantations is not included as farm activities.

Overall, there are 362 million 'main' workers and 119 million 'marginal' according to the Census of India 2011.

Economic Census. The economic census is a complete enumeration of non-agricultural enterprises in India. While recent economic censuses have expanded the scope to cover establishments engaged in various agricultural activities, the strength of the economic census lies in providing firm-level information on employment for non-agricultural establishment.[29] In 2012-13 there were approx. 45 million non-agricultural enterprises, employing 108 million workers in the country. An advantage of the economic census is that it allows us to explore heterogeneous impacts on firms by their size and disaggregate the specific sub-sectors which is not possible in the Census of India.

Irrigation. Dams, especially embankment dams, are an an important source of irrigation in India. The mean (median) number of dams in an Indian district has increased from 2.05 (0) to 7.84 (1) in the period 1970 to 1999. Although there has been a significant rise in the number of dams over the years, their distribution is not uniform across states. Instead, the new dams have been primarily concentrated in the western region, especially Maharashtra and Gujrat (Duflo and Pande, 2007).

Embankment dams are built using an artificial wall dividing the area into *catchment* and *command* areas. *Catchment* area refers to upstream part of the dam from which the water flows in, whereas *command* areas refers to the downstream part from where the water is then channelled for irrigation through a network of canals. By design, the benefits of these dams for irrigation purposes are limited to those who live in the *command* area.

In India, constructing a dam requires approval both by state and national governments, and is thus subject to a proper cost benefit analysis (Asmal et al., 2000). Although the benefit is often measured in terms of agricultural output and the value of power to be generated, the costs are much more complicated to evaluate (Duflo and Pande, 2007). Geography is an important

[29] Public administration, defence and social security activities are excluded

determinant of the cost: for example, a river that flows at a moderate incline makes it easier and cheaper to construct a dam. Additional hidden costs includes dam's impact on land productivity due to water-logging and water salinity, and the concomitant impacts on the health of those living in nearby areas, and displacement of the people to name a few.

This form of irrigation using canals connected to dams is the most important form of irrigation in India because it is cheaper than other alternatives. Ground water and small dykes are two potential alternatives. In contrast to dams, these alternative are less effective, especially in areas like India with high seasonal rainfall (Biswas and Tortajada, 2001).

Towns. An important element of our analysis is the differential effects of being in a command area on villages and towns. It is therefore important to clearly articulate the administrative, demographic, and economic characteristics that distinguish towns from villages.

Administratively, towns differ from villages due to their being governed by municipal corporations, and municipality and city councils. In contrast, most village-level administration is undertaken by larger administrative units, such as sub-district and district authorities, though some functions are reserved to village governing bodies (i.e., panchayats). This administrative distinction has one notable exception, however: specifically, the "census towns," which are classified by the Registrar General of India as being towns due to their population size, density, and labor force composition,[30] but which have not yet been granted official statutory township status by the government (Pradhan, 2017). Despite lacking urban governing institutions, census towns display similar levels of prosperity and economic diversification as statutory towns, and differ markedly from the typical village. We, therefore, refer to all the non-village sites as "towns."

Demographically, towns have far larger populations than villages, and a far higher population density. Economically, towns differ from villages by the share of agricultural in local employment and production. Though towns may include significant agricultural activities, and to employ a substantial share of the land within their boundaries to agriculture, the scale of these activities is

[30] Census towns must meet three criteria: (1) a population greater than 5000; (2) a population density above 400 individuals per square kilometer; and (3) a male labor force which is less than 25% engaged in agriculture.

vastly smaller. In addition, towns feature a far larger number of firms engaged in manufacturing and other non-agricultural activities, which are more likely to be formalized, and operate on a far larger scale. Towns also feature substantial retail, wholesale, and transport sectors: indeed, one of the major drivers of the recent growth of towns in rural areas has been their role as markets and distribution centers for nearby villages.

Appendix Table A1 depicts some of the key differences between villages and towns. In column (1) of Panel B are given the mean characteristics of villages in the study area. In column (2) we present the difference between towns and villages; in column (3) we include project-area fixed effects; and in column (4) we restrict the sample of towns to those occupying less than 30 square kilometers. This table highlights the starkly different character of towns and villages, with the latter having more built-up area and less agricultural, more light density, higher population and population density, and greater asset holdings and household amenities. Towns also feature less agricultural employment, and more employment in service and manufacturing firms, of which a larger share is in large firms with more than 10 or 100 workers.

In Appendix Figure A2 we show the relationship between distance to nearest town and the structure of the economy (with towns taking a distance value of 0), again illustrating both the differences between villages and towns, and the somewhat more "urban" economic structure of those villages located in the immediate vicinity of towns.

Town Formation. One potential concern with our analysis is that whether a village graduates to the status of being a town is itself endogenous. This could lead to the result that a village experiencing a large increase in population could be reclassified as a town with a small population. In Appendix Table A2 we test for endogenous town formation.

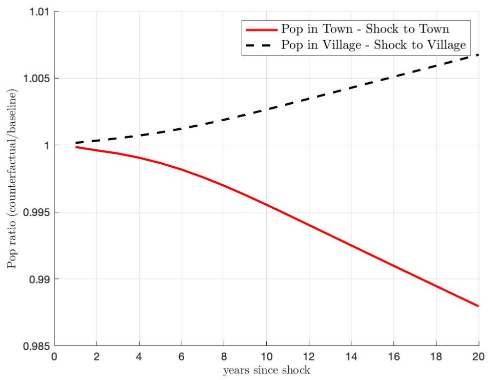
According to the Census of India, a 'census town' is defined as one where the population exceeds 5,000, population density is more than 400 persons per sq km, and more than 75 percent of main male working population is employed outside the agricultural sector. We therefore define a 'marginal sample' as villages and towns that were close to meeting the criterion i.e. a population between 4,000 and 6,000 people, a population density of more than 350 persons per km sq, and male labor force greater than 70 percent that is engaged in non-agricultural production.

In Appendix Table A2, column (1), we restrict the sample to villages and towns that were close to meeting the criterion for township formation (the 'marginal sample') and estimate the impact of being in the program areas on attaining township status. In columns (2) and (3), we restrict the sample to all towns, and take as the outcome variable an indicator for whether the town already had township status in 1951 and 1971, respectively. In columns (4)–(5), we take as the outcome the log *area* within towns, where land within villages take a value of 0, and towns take the natural log of their areas.

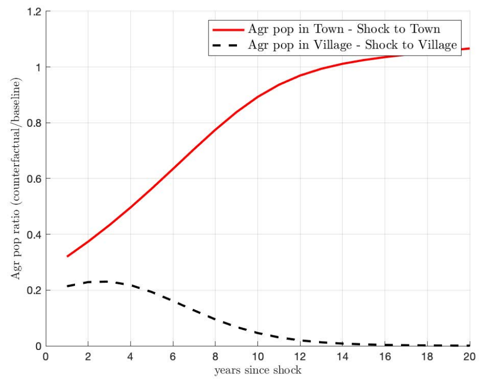
We find no impact on whether villages are graduated to township status, nor on how early existing towns were formed. In addition, we find no impact on the total area of the command area that is within a township (column 4), or on the size of towns (column 5). Essentially, this means that the 6.3% increase in village population in treatment areas (coupled with absence of any change in the labor share in agriculture) was insufficiently large to graduate villages to township status. This is intuitive, given the dramatically larger populations of towns, and their far smaller agricultural labor shares. The area covered by towns was no different in control and treatment areas: towns were simply less populated and built-up in the latter.

A3 Appendix: Additional Figures and Tables

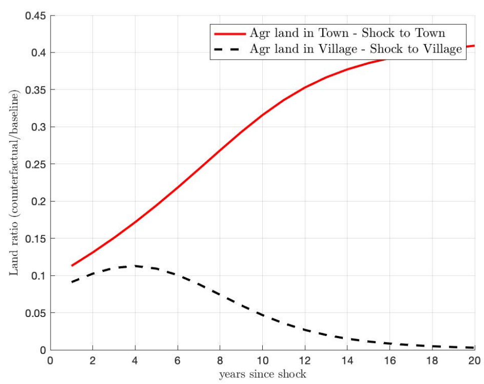
Figure A1: Numerical Examples of the Impact of an Agricultural Productivity Shock in a Town versus a Village



A1.1: Population



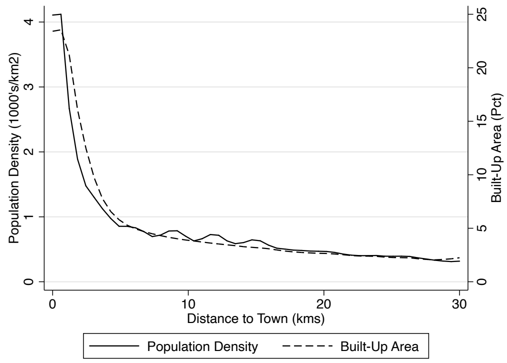
A1.2: Share of Agricultural Workers



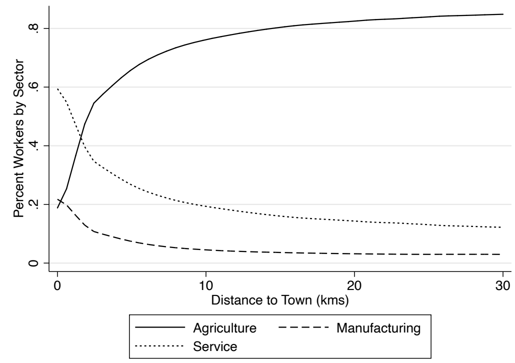
A1.3: Share of Agricultural Land

Notes: These figures show a numerical example of the impact of an agricultural productivity shock in a village versus a town using our stylized spatial economy model. In black dashed line we have the path for villages (for an agricultural productivity shock in villages) and in red line the path for towns (for an agricultural productivity shock in towns). It shows that the shock has a positive effect on population in village, but a negative in town. In addition, the figure shows that the shock, in both cases, has a positive effect on the share of labor and land in agriculture, but the effect tends to be larger in towns.

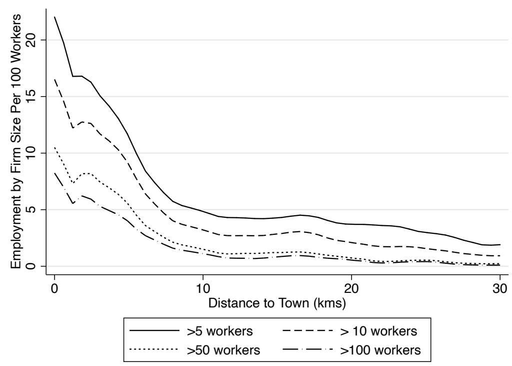
Figure A2: Spatial Distribution of Economic Activity



A2.1: Urbanization



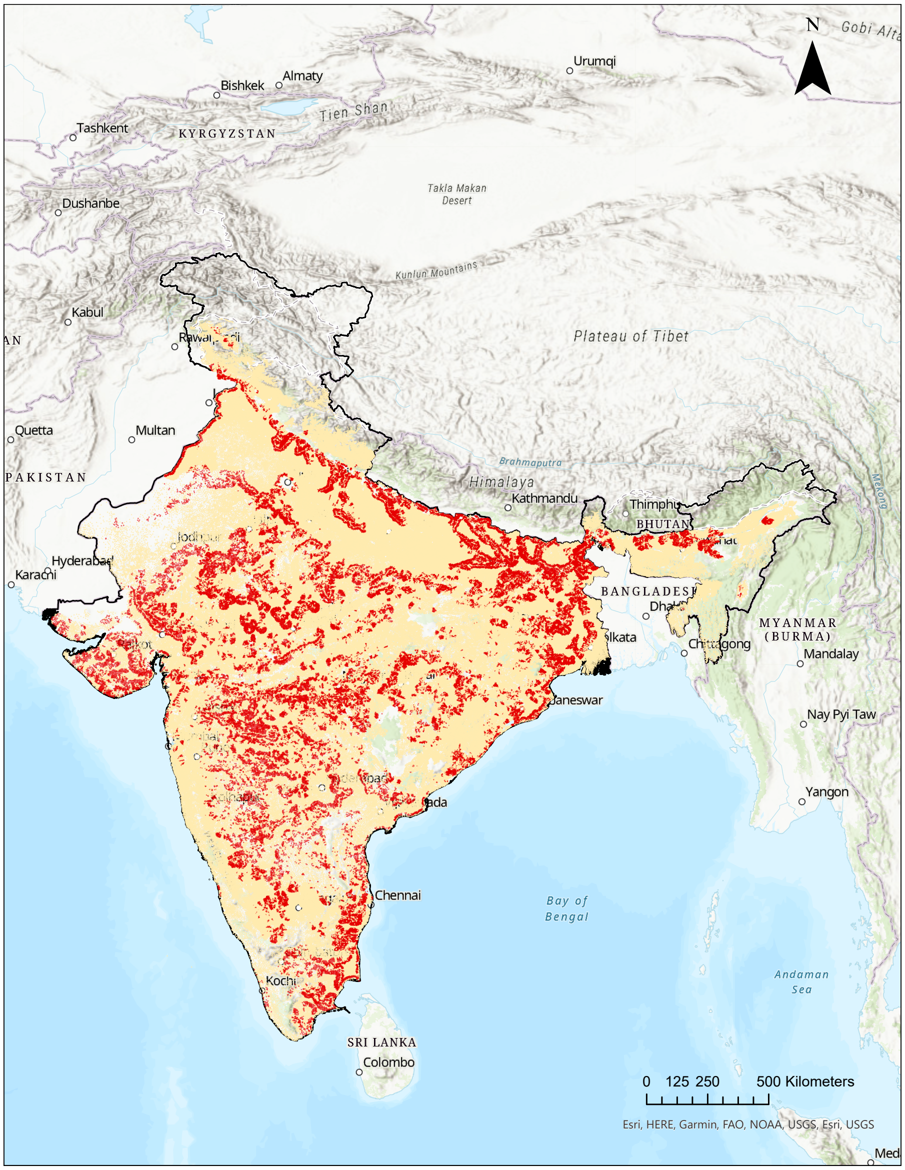
A2.2: Labor force



A2.3: Firm Size

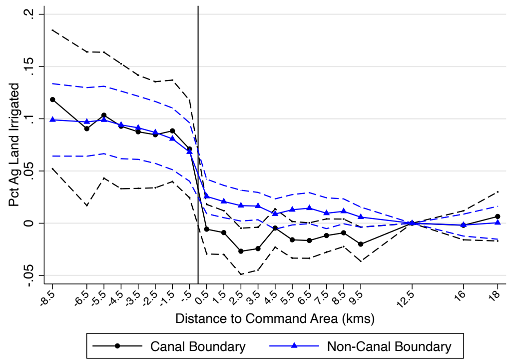
Notes: The above figure plots the spatial distribution of economic activity of villages relative to towns. Distance from village centroid to the nearest town (in km) is on the x-axis. Figure A2.1 depicts the population density per 1,000 square km (Census of India 2011) on the left y-axis and percentage of built-up area on the right y-axis. Figure A2.2 depicts percent of workers in agriculture (Census of India 2011), manufacturing and service sectors (Economic Census 2012-13). Figure A2.3 depicts employment by firm size (Economic Census 2012-13).

Figure A3: Villages in the trimmed sample

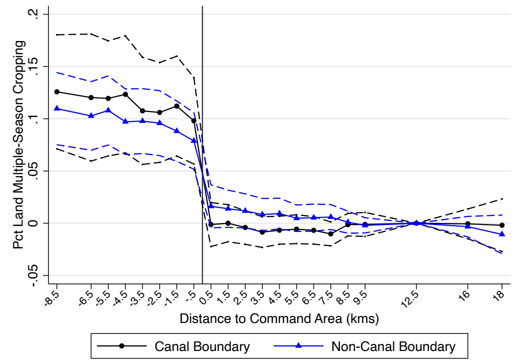


Notes: Each dot corresponds to a village in India. The villages in the study sample are denoted by a red dot, while those not in the study sample are denoted in topaz sand color. There are approx 145,000 villages in the trimmed sample, which account for 22 percent of the nearly 650,000 villages in the country.

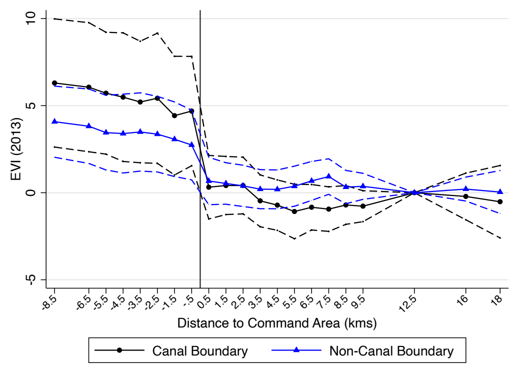
Figure A4: Agriculture by Boundary Type



A4.1: Pct of Agriculture Area Irrigated



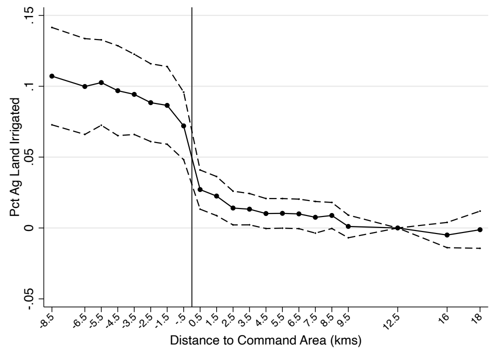
A4.2: Multi-Season Cropping



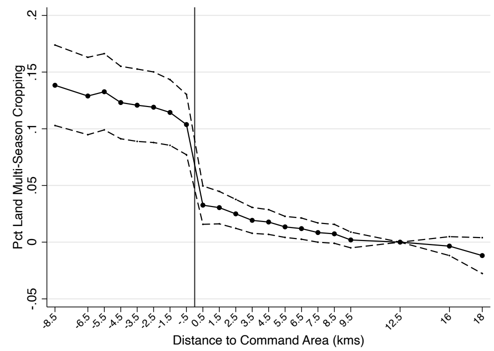
A4.3: Dry Season Vegetation

Notes: The above figure compares key agricultural outcomes in villages inside the command area (to the left of 0) with those just outside (to the right of 0). Distance from village centroid to the command area (in km) is on the x-axis. The solid line represents results from a regression of outcomes on canal command area treatment dummy, binned distances, controls and 5 km boundary segment fixed effects. Standard errors are clustered at the project code level. The dotted lines represent 95% confidence intervals. The black lines refers to the sample which only had a contiguous canal and command area boundary, while the blue lines refers to the non-canal boundary (see definitions in text). Figure A4.1 depicts area under irrigation in percent of cultivable land (Census of India 2011), Figure A4.2 depicts land area that is cropped twice or thrice in percent of agricultural area (NRSC/ISRO 2011-12), and Figure A4.3 depicts dry season vegetation indices in percent of total village area (MODIS EVI 20013).

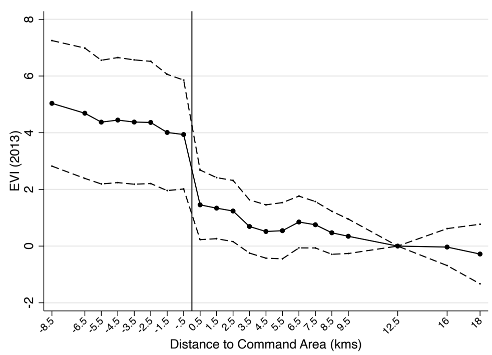
Figure A5: Agriculture and Development, w/o Geographic Controls



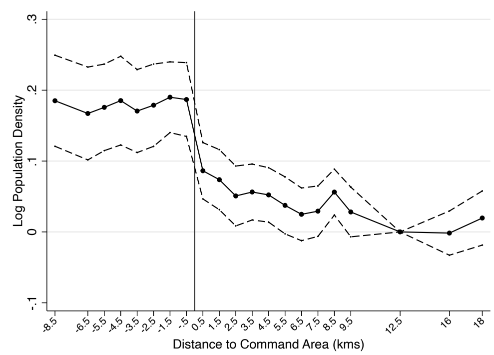
A5.1: Pct of Agriculture Area Irrigated



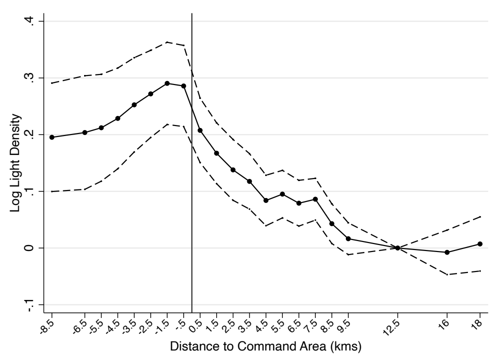
A5.2: Multi-Season Cropping



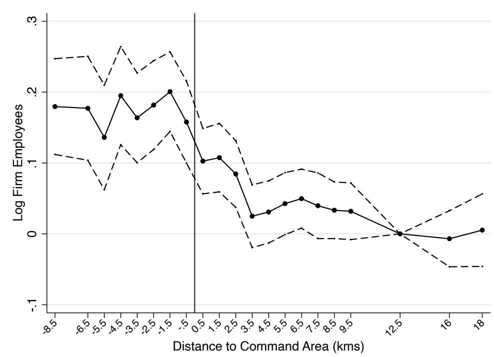
A5.3: Dry Season Vegetation



A5.4: Log Population Density



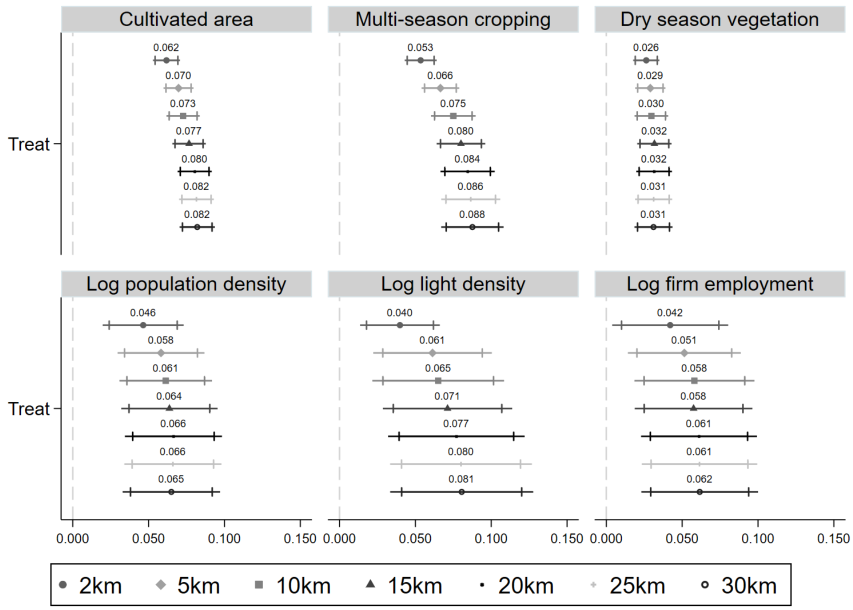
A5.5: Log Light Density



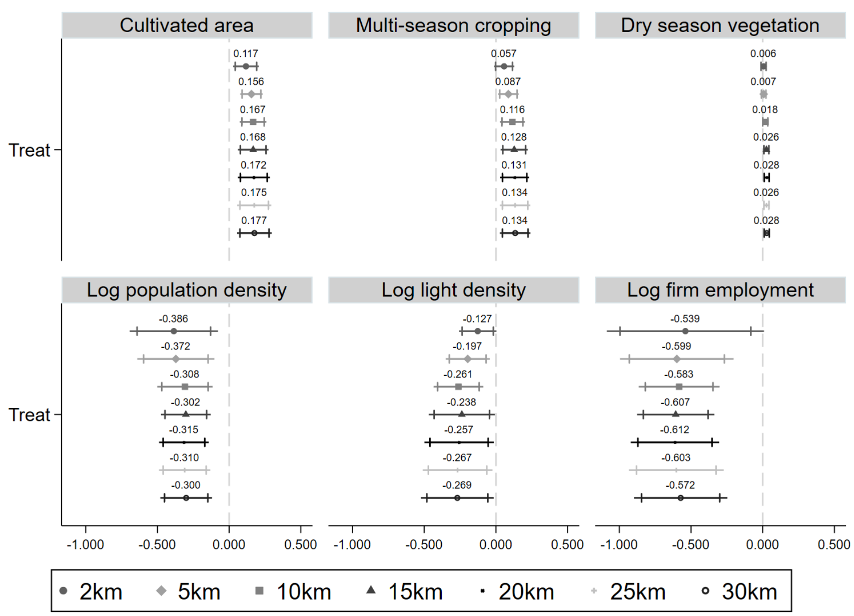
A5.6: Log Firm Employment

Notes: The above figure compares key agricultural and development outcomes in villages inside the command area (to the left of 0) with those just outside (to the right of 0). Distance to the command area (in km) is on the x-axis. The solid line represents results from a regression of outcomes on canal command area treatment dummy, binned distances, controls and 5 km boundary segment fixed effects. Standard errors are clustered at the project code level. The dotted lines represent 95% confidence intervals. Figure A5.1 depicts area under irrigation as percent of cultivable land; Figure A5.2 depicts land area that is cropped twice or thrice as percentage of agricultural area; and Figure A5.3 depicts dry season vegetation indices as percentage of total village area. Figure A5.4 depicts log population density; Figure A5.5 depicts mean nighttime lights per sq km; Figure A5.6 depicts number of employees in firms across manufacturing, agriculture and services enterprises.

Figure A6: Robustness to varying bandwidths



A6.1: Villages



A6.2: Towns

Notes: The figures plot the impact on key agricultural and non-agricultural outcomes for villages and towns using alternative bandwidths (2 km, 5 km, 10 km, 15 km, 20 km, 25 km and 30 km). Capped spike intervals report the 90 percent while the longer intervals report the 95 percent confidence intervals. Agricultural outcomes are derived from satellite data. Cultivated area refers to percentage of area cultivated; multi-season cropping refers to area cropped twice or thrice in a year; and dry-season vegetation refer to MODIS EVI 2013. The non-agricultural outcomes are: population density; night light density; and built-up area.

Figure A7: Town Presence

Notes: This figure plots the number of towns (demeaned by the number in 1951) against year. The sample is disaggregated into ”in sample” towns, which are those located in the control and treatment groups from the study sample; and ”out of sample” towns, which are towns located outside of the study sample (excluding out-of-sample towns located within command areas).

Table A1: Summary Statistics

Num Command Areas			1,533	
Median Year Completion			1977	
Num Villages inside Command Area			245,131	
Num Towns inside Command Area			2,879	
Num Villages inside Command Area (in Study Sample)			73,817	
Num Towns inside Command Area (in Study Sample)			886	

	Village Mean	Town – Village Mean		
	(1)	(2)	(3)	(4)
Total Area (km2)	4.077	9.388***	8.298***	3.577***
		(1.037)	(0.884)	(0.279)
Share Area Built-Up	0.050	0.193***	0.179***	0.191***
		(0.009)	(0.008)	(0.008)
Share Area Agriculture	0.625	-0.243***	-0.176***	-0.191***
		(0.020)	(0.021)	(0.025)
Light Density	6.075	19.437***	16.393***	16.098***
		(1.077)	(0.962)	(1.049)
Tot Population (1,000s)	1.618	39.805***	39.509***	24.654***
		(3.025)	(3.006)	(1.294)
Population Density (1,000s/km2)	0.712	3.326***	3.422***	3.543***
		(0.181)	(0.163)	(0.172)
Pct Male Workers Ag	0.757	-0.589***	-0.513***	-0.510***
		(0.009)	(0.010)	(0.011)
Employees in Firms (100s)	1.365	66.561***	64.874***	41.305***
		(5.484)	(5.336)	(2.858)
Employees in Manu Firms (100s)	0.291	18.429***	17.938***	11.864***
		(1.907)	(1.909)	(1.100)
Share Employees in Firms $>$10 Workers	0.060	0.114***	0.083***	0.078***
		(0.008)	(0.007)	(0.008)
Share Employees in Firms $>$100 Workers	0.007	0.044***	0.036***	0.033***
		(0.004)	(0.003)	(0.003)
Pct HHs w/TV	0.282	0.367***	0.218***	0.215***
		(0.015)	(0.011)	(0.011)
Pct HHs w/Telephone	0.522	0.204***	0.157***	0.157***
		(0.013)	(0.008)	(0.008)
Pct HHs w/Scooter	0.143	0.135***	0.088***	0.085***
		(0.013)	(0.006)	(0.006)
Pct HHs w/Brick Wall	0.473	0.268***	0.204***	0.207***
		(0.014)	(0.017)	(0.018)
Pct HHs w/Water Source on Premises	0.321	0.260***	0.228***	0.224***
		(0.024)	(0.013)	(0.014)
Project Area F.E.s			Yes	Yes
Area $<$30 km2				Yes

Note: This table reports descriptive statistics for the estimating sample. The first panel reports basic information on the coverage of the irrigation projects. The second panel reports the mean of various outcome variables by treatment status. Column (1) reports the mean for villages and columns (2)-(4) report the mean difference between towns and villages. Column (2) reports the unconditional mean, column (3) adds project fixed effects and column (4) restricts the sample to towns with areas smaller than 30 sq km.

Table A2: Town Formation

	Town	Existing Town 1951	Existing Town 1971	Log Town Area	
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.019	0.016	-0.006	-0.015	-0.024
	(0.054)	(0.030)	(0.032)	(0.010)	(0.060)
R-squared	0.339	0.456	0.382	0.104	0.507
N	430	1546	1546	147885	1546
Town Sample	Yes	Yes	Yes	Yes	Yes
Village Sample	Yes			Yes	
Marginal Sample	Yes				

Note: In column (1), the outcome variable is an indicator for township status; and the sample is restricted to villages and towns that have a population between 4000–6000, a male labor force share $<$0.30 in agriculture, and a population density greater than 350 per square km. In columns (2) and (3), the outcome variable is an indicator for having been a township in 1951 and 1971, respectively. Columns (4) and (5) take as the outcome the log area of towns, which takes a value of 0 for villages.

Table A3: Balance, Geographic Features

	Control Mean	Difference: Full Sample	Difference: Trimmed Sample	
				RD
	(1)	(2)	(3)	(4)
Altitude	202.468	-21.209***	-5.706***	-0.915*
		(1.591)	(0.599)	(0.544)
		[-0.056]	[-0.015]	[-0.002]
Ruggedness Index	38.796	-13.029***	-2.148***	0.255
		(0.974)	(0.234)	(0.242)
		[-0.109]	[-0.018]	[0.002]
Distance Major River	30.887	-0.063	0.444	0.423**
		(0.211)	(0.277)	(0.166)
		[-0.001]	[0.009]	[0.009]
Alluvial Aquifer	0.556	0.047***	0.023***	0.015*
		(0.007)	(0.006)	(0.008)
		[0.094]	[0.046]	[0.030]

Notes: Table reports results from equation: $y_{idb} = \alpha + \beta C_i + \nu_d + \mu_b + \varepsilon_{idb}$ where, y_{idb} is an outcome of interest in village i in district d in a 10 km buffer around boundary segments b; C_i is an indicator variable indicating whether the centroid of a village is located inside command area or not; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. The outcomes are altitude (in meters), terrain ruggedness index derived from USGS digital elevation models, distance to river (in kms), and whether a village lies on top of an alluvium/water-deposited aquifer. Standardized z-scores for the outcomes are in square brackets. Column 1 reports the mean of the outcome outside the command area; Column 2 reports the difference between villages inside and outside the command area in the full sample; Column 3 and Column 4 refer to the trimmed sample. (In the trimmed sample, the sample is restricted to villages for which the average slope on both sides is less than 1.5 degrees; boundaries where the canal is within 500m of a river are also excluded.) Column 3 uses the baseline specification mentioned above; Column 4 additionally includes treatment-interacted control for distance to the command area boundary. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A4: Treatment Effects at Canal Boundaries

	Pct				Log			Pct HHs/w	
			Area						
	Ag Area Canal	Ag Area Irrigated	Multi-Season Cropping	EVI (2013)	Population Density	Light Density	Firm Employment	TV	Brick Wall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Canal Boundary									
Treatment	0.100***	0.069***	0.091***	4.304***	0.094***	0.006	0.056*	0.009*	0.010*
	(0.013)	(0.012)	(0.013)	(0.836)	(0.025)	(0.032)	(0.033)	(0.005)	(0.006)
R-squared	0.610	0.693	0.735	0.845	0.459	0.779	0.611	0.765	0.641
N	29980	30004	30405	28476	28549	28805	27305	28454	28454
Panel B: Canal Boundary RD Specification									
Treatment	0.133***	0.089***	0.106***	4.174***	0.102**	0.018	0.032	0.016*	0.027***
	(0.022)	(0.019)	(0.018)	(1.345)	(0.041)	(0.051)	(0.051)	(0.008)	(0.009)
R-squared	0.648	0.706	0.741	0.850	0.467	0.779	0.606	0.765	0.643
N	23220	23238	23613	21855	21964	22128	20919	21912	21912

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i (village or town) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. In addition to the usual sample restrictions (locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded), the estimating sample comprises only of villages for which the nearest command area boundary coincides with a canal. Column 1 reports area irrigated using canals (as percentage of cultivated area), Column 2 area irrigated using any source (as percentage of cultivated area), Column 3 area cropped twice or thrice in a year (from NSRC/ISRO), Column 4 dry season vegetation (from MODIS EVI 2013), Column 5 ln(population density) from Census of India 2011, Column 6 ln(light density), Column 7 ln(total employed workers) in all enterprises from Economic Census 2012-13, and Column 8 and Column 9 report percentage of households who own a TV or have a brick wall respectively. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A5: Placebo Regressions, 1991

	Pct Ag Land Irrigated		Log				
	Canal	Any	Pop Density	Light Density	Pct Workers	Pct Worker Ag	Log Employees
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.002	0.014	-0.007	-0.014	0.001	0.009	0.019
	(0.002)	(0.009)	(0.028)	(0.032)	(0.003)	(0.005)	(0.054)
Control Mean	0.015	0.200	0.655	0.788	0.411	0.860	2.656
R-squared	0.565	0.763	0.463	0.703	0.572	0.437	0.543
N	13447	13447	12715	12828	12777	12774	11188

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in the year 1991 in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. In addition to the usual sample restrictions (locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded), the the estimating sample comprises only of command areas/irrigation projects that started after 1991. Column 1 reports area irrigated using canals (as percentage of cultivated area), Column 2 area irrigated using any source (as percentage of cultivated area), Column 3 ln(population density), Column 4 ln(mean light density), Column 5 total employment (as percentage of population), Column 6 is farmers (as a percentage of workers) where farmers = own-farm workers/cultivators + ag laborers, and Column 7 is ln(total employed workers) in all enterprises. Data from Column 1-3 and Column 5-6 comes from Census of India 1991; Column 4 is from NOAA 1993; Column 7 is from Economic Census 1991. Only limited outcomes are shown because remote sensing data from NSRC/ISRO and MODIS EVI are not available for 1991. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A6: Robustness tests using boundary distance controls

	Pct				Log			Pct HHs/w	
			Area						
	Ag Area Canal	Ag Area Irrigated	Multi-Season Cropping	EVI (2013)	Population Density	Light Density	Firm Employment	TV	Brick Wall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Villages									
Treatment	0.090***	0.055***	0.076***	2.785***	0.062***	0.090***	0.051*	0.008**	0.014**
	(0.011)	(0.009)	(0.009)	(0.807)	(0.023)	(0.034)	(0.031)	(0.004)	(0.007)
R-squared	0.378	0.687	0.723	0.836	0.488	0.742	0.533	0.754	0.695
N	115741	115835	117349	99141	108426	104474	101102	108051	108051
Panel B: Towns									
Treatment	NA	NA	0.130***	-0.048	-0.430*	-0.353***	-0.782***	-0.039	-0.087**
			(0.049)	(1.531)	(0.222)	(0.111)	(0.217)	(0.042)	(0.038)
R-squared			0.717	0.791	0.616	0.857	0.613	0.859	0.735
N			555	534	564	548	564	564	564

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i\Gamma + \lambda Distance_i + \psi(C_i \times Distance_i) + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i (villages in Panel A or towns in Panel B) in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; $Distance_i$ is distance to the command area boundary; $C_i \times Distance_i$ is the interaction treatment status and distance to boundary; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded; and villages that are partially inside the command area are also omitted. Column 1 reports area irrigated using canals (as percentage of cultivated area), Column 2 area irrigated using any source (as percentage of cultivated area), Column 3 are cropped twice or thrice in a year (from NSRC/ISRO), Column 4 reports dry season vegetation (from MODIS EVI 2013), Column 5 ln(population density) from Census of India 2011, Column 6 ln(mean nighttime luminosity score per sq km) from NOAA 2013, Column 7 ln(total employed workers) in all enterprises from Economic Census 2012-13, Column 8 and Column 9 report percentage of households who own a TV or have a brick wall respectively. "NA" denotes that data is not available. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A7: Treatment Effects excluding Villages <2 kms Inside Boundary

	Pct				Log			Pct HHs/w	
			Area						
	Ag Area Canal	Ag Area Irrigated	Multi-Season Cropping	EVI (2013)	Population Density	Light Density	Firm Employment	TV	Brick Wall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.112***	0.073***	0.091***	3.372***	0.053**	0.049	0.060**	0.009**	0.018**
	(0.011)	(0.009)	(0.010)	(0.800)	(0.023)	(0.034)	(0.028)	(0.004)	(0.008)
R-squared	0.368	0.689	0.725	0.832	0.496	0.742	0.545	0.761	0.697
N	118685	118783	120393	103647	111892	108859	104808	111431	111431

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. In addition to the usual sample restrictions (locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded), the estimating sample drops villages inside the irrigation project which are within a 2 km distance from the command area border. Column 1 reports area irrigated using canals (as percentage of cultivated area), Column 2 area irrigated using any source (as percentage of cultivated area), Column 3 are cropped twice or thrice in a year (from NSRC/ISRO), Column 4 reports dry season vegetation (from MODIS EVI 2013), Column 5 ln(population density) from Census of India 2011, Column 6 ln(light density), Column 7 ln(total employed workers) in all enterprises from Economic Census 2012-13, Column 8 and Column 9 report percentage of households who own a TV or have a brick wall respectively. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A8: Robustness Tests

	Pct				Log			Pct HHs/w	
			Area						
	Ag Area Canal	Ag Area Irrigated	Multi-Season Cropping	EVI (2013)	Population Density	Light Density	Firm Employment	TV	Brick Wall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Dropping Boundary Villages									
Treatment	0.108***	0.067***	0.086***	3.180***	0.061**	0.086**	0.050	0.011***	0.021***
	(0.011)	(0.010)	(0.011)	(0.860)	(0.024)	(0.034)	(0.031)	(0.004)	(0.008)
R-squared	0.377	0.687	0.723	0.835	0.488	0.741	0.533	0.754	0.695
N	115741	115835	117349	99141	108426	104474	101102	108051	108051
Panel B: Winsorized									
Treatment	0.079***	0.056***	0.073***	2.834***	0.058***	0.071***	0.057***	0.009***	0.015***
	(0.007)	(0.007)	(0.007)	(0.547)	(0.014)	(0.021)	(0.020)	(0.003)	(0.005)
R-squared	0.512	0.679	0.720	0.831	0.534	0.740	0.539	0.754	0.692
N	142628	142736	144380	124265	134305	130487	125796	133720	133720
Panel C: Conley Errors									
Treatment	0.091***	0.066***	0.082***	3.384***	0.078***	0.087***	0.072***	0.011***	0.019***
	(0.009)	(0.010)	(0.011)	(0.815)	(0.012)	(0.022)	(0.021)	(0.003)	(0.005)
R-squared	0.017	0.019	0.069	0.025	0.053	0.009	0.183	0.006	0.005
N	145163	145269	146803	126845	137315	131604	128912	136694	136694

Notes: Table reports results from: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ where, y_{ipdb} is an outcome of interest in location i in a 10 km buffer around irrigation project p in district d along boundary segment b; C_i is an indicator variable for whether the centroid of a location lies inside a command area of project p or not; X_i is a vector of geographic characteristics like altitude, ruggedness, distance to major river, type of groundwater aquifer underlying the location, the (log) area of the location; ν_d are district fixed effects; and μ_b are 5 km boundary segment fixed effects. Estimating sample is restricted to locations for which the average slope on both sides of the boundary is less than 1.5 degrees and to locations with area less than 30 sq. km; boundaries where the canal is within 500m of a river are also excluded. In Panel A, the estimating sample is restricted further and villages whose boundary overlaps with the command area border are dropped; panel B winsorizes the outcomes at the 5% and 95% level; panel C reports the Conley adjusted standard errors (300 km radius). Column 1 reports area irrigated using canals (as percentage of cultivated area), Column 2 area irrigated using any source (as percentage of cultivated area), Column 3 are cropped twice or thrice in a year (from NSRC/ISRO), Column 4 reports dry season vegetation (from MODIS EVI 2013), Column 5 ln(population density) from Census of India 2011, Column 6 ln(light density), Column 7 ln(total employed workers) in all enterprises from Economic Census 2012-13, Column 8 and Column 9 report percentage of households who own a TV or have a brick wall respectively. Standard errors are clustered by command area (irrigation project) to conservatively account for potential spatial correlation * p<0.10, ** p<0.05, *** p<0.01.