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

David Blakeslee[†] Aaditya Dar[‡] Ram Fishman[§]

Samreen Malik[¶] Heitor Pelegrina[‖] Karan Singh[**]

October 2022

Abstract

We study the long-term impact of large-scale irrigation infrastructure on the composition of local economic activity in India. Our analysis uses high-resolution spatial data covering approximately 150,000 villages and towns and exploits spatial discontinuities in the coverage of irrigation projects. Irrigation increases agricultural output, wealth, and population density in rural villages. However, in towns it reduces population and nightlight density, the size of the non-agricultural sector, and large-firm activity. These results highlight the heterogeneous impacts that agricultural productivity gains can have on the patterns local economic development.

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

[*] We are grateful to the 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 the University of Manchester, UK for helpful discussion and feedback. We would also like to thank Vedarshi Shastry and Chanchal Kumar Singh for their 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 long-term agricultural productivity shocks on local economic development in India. Access to irrigation is considered to be key to agricultural development, and the establishment of irrigation infrastructure has long been one of the leading forms of public investment in the sector in much of the developing world, and particularly so 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. Our analysis reveals that this widespread transformation had heterogeneous impacts on local economic development.

On the one hand, we provide evidence that these irrigation projects had a positive impact on agricultural productivity in villages by allowing them to expand crop production to seasons when it had previously been nonviable. We furthermore show that these areas experienced increases in population density and indicators of economic development (assets and nightlights), but there were no significant changes in the share of agricultural workers and only a modest increase in small-firm activity.

The effects of irrigation in towns, however, were notably different. Towns differ from villages in being larger, more densely populated, and having economies oriented towards non-agricultural production and trade. We show that towns located within project areas experienced declines in population density and indicators of economic development. Importantly, towns also experienced a substantial decline in the scale of manufacturing activity and the presence of large firms, as well as a shift in the labor force away from non-agricultural employment.

Our analysis is guided by a stylized, spatial economy model with endogenous manufacturing productivity growth that illustrates the effects of an agricultural productivity shock on the spatial and sectoral allocation of workers. The model indicates that these effects depend substantially on the type of region that is hit by the shock. A positive, permanent agricultural productivity shock slows down productivity growth in the manufacturing sector, which in the long-run can generate a reduction in population and real wages relative to a scenario with no shock, akin to Matsuyama (1992). The extent to which this long-run mechanism dominates the short-run benefits of the agricultural productivity gain depends on the size of the manufacturing sector in a region, and the degree of labor mobility between regions of the country. As a result of the shock, regions that are primarily agricultural tend to have an increase in population in the long-run despite the slower manufacturing productivity growth, whereas more urbanized regions tend to experience the opposite effect.

We make use of fine spatial data on more than 1,500 major surface irrigation projects in India, 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 (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 locations proximate to one another but on opposite sides of the boundary, while controlling for geographic features and imposing sampling restrictions to ensure comparability.

Our paper joins a growing literature on the causal impact of different forms of agricultural productivity gains on the broader economy, both across (McArthur and McCord, 2017; Gollin et al., 2021) and within (Hornbeck and Keskin, 2015; Bustos et al., 2016, 2020) countries. In the context of India, Foster and Rosenzweig (2004) show that high rates of crop yield growth in India are correlated with lower industrial growth across a nationally representative sample of villages. While existing studies conduct their analysis at relatively high levels of administrative aggregation, we contribute to the literature by using higher resolution data to show that impacts can dramatically vary in space and by baseline levels of urbanization.

We also contribute to the literature on the impacts of large irrigation infrastructure. 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 (Hansen et al., 2011; Strobl and Strobl, 2011; Blanc and Strobl, 2014; Olmstead and Sigman, 2015; Jones et al., 2022; Dillon and Fishman, 2019; Zaveri et al., 2020).[1] These studies have documented important effects of irrigation on agricultural output, income volatility, and poverty rates, but have not investigated impacts on non-agricultural economic activity, which is a primary focus of this paper.[2]

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 spatial RDD, with elevation relative to the canal as the running variable, and similarly find positive impacts on agricultural productivity, population density, and measures of economic development in rural areas.[4]

The key distinction between the approach of these other papers and our own is in the treatment of towns. Boudot-Reddy and Butler (2021) focus exclusively on villages; while Asher et al. (2021) do not distinguish between towns and villages in their RDD analysis, yielding results that are dominated by village effects, and provide a separate analysis estimating impacts on the India-

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

[2] 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).

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

[4] We note that relative elevation is largely a monotonic function of the distance from the command area boundary, the running variable used in our paper (see Figure 2.1).

wide distribution of town location and growth.[5] In contrast, we show, both theoretically and empirically, that the effects of irrigation on the economic development of villages and towns are strikingly different.

In the next section, we present a stylized spatial economy model to derive predictions for our empirical analysis. Data and summary statistics are discussed in Section 3, followed by the empirical strategy in Section 4. Finally, we present our results in Section 5 and conclude the paper in Section 6.

2 A Stylized Model

This section develops a stylized spatial economy model, which we use to illustrate the potential effects of an agricultural productivity shock on the spatial and sectoral allocation of workers. We show that these effects depend fundamentally on the type of region that is hit by the shock. Specifically, we consider three types of regions: (i) a rural location that specializes in agriculture and releases workers to the rest of the country; (ii) an urban location that specializes in manufacturing and absorbs workers from the rest of the country; and (iii) a region which is composed of multiple types of locations. Later, in the context of India, we organize our empirical analysis around these three types of geographic units.

Below, we present the general setup of our model, we then turn to the effects of agricultural productivity shocks in each type of geographic region. To minimize notation, this section focuses on the main results and intuition, relegating a full description of the model to Appendix A1.

2.1 Setup and Equilibrium

Consider a location ℓ^o, which we treat as a small, open economy within a country. The economy operates over discrete time. There are two sectors, agriculture (A) and manufacturing (M). We think of the manufacturing sector as capturing the production of tradable goods produced by

[5] The latter is based on a difference-in-differences analysis comparing areas within and near the command areas to more distant areas.

larger firms.[6] Preferences are Cobb-Douglas between goods produced by each sector. There is a population of workers N, which can choose, in every period, whether to work in location ℓ^o or in a location ℓ^* that represents the rest of the country. This choice is based on the real wage of each location and on a location-specific taste parameter that is heterogeneous across workers.[7] There is also a mass of land L in ℓ^o, which is heterogeneous in terms of a sector-specific adjustment cost. Landowners allocate land to sectors based on this sector-specific adjustment costs and the land rents obtained in each sector.[8] Markets are perfectly competitive.[9]

The technology to produce goods, $q_{k,t}(\ell^o)$, is given by:

$$q_{k,t}(\ell^o) = A_{k,t}(\ell^o)\left(L_{k,t}(\ell^o)\right)^{\alpha_k}\left(N_{k,t}(\ell^o)\right)^{1-\alpha_k} \tag{1}$$

where $A_{k,t}(\ell^o)$ is the productivity, $L_{k,t}(\ell^o)$ the use of land, $N_{k,t}(\ell^o)$ the use of labor, α_k is the share of land in production, and t and k index the period and the sector, respectively. Agricultural productivity is fixed at its initial value, $A_{A,t}(\ell^o) = A_{A,0}(\ell^o)$. Akin to Matsuyama (1992), manufacturing productivity evolves endogenously, according to:

$$A_{M,t+1}(\ell^o) = A_{M,t}(\ell^o) + \gamma\left[N_{M,t}(\ell^o)/N_t(\ell^o)\right] \tag{2}$$

where $\gamma > 0$ is a parameter controlling the speed of productivity growth and $N_t(\ell^o)$ is the mass of workers living in ℓ^o. This expression captures, in a simple form, local forms of productivity growth that takes place against the backdrop of global, economy-wide productivity growth.[10]

[6] To keep matters simple, we do not introduce a service, non-tradable sector in the model, which could capture smaller shops and firms oriented towards local consumers.

[7] Appendix A1 provides a simplified formulation in which we do not incorporate amenities or agglomeration economies. However, the structure of the model is sufficiently flexible and can accommodate different forms of agglomeration economies using recent approaches in the spatial economy literature. We also note that the model is flexible in terms of its geographic structure and that, in principle, it could be extended to incorporate multiple outside options for workers.

[8] Appendix A1 solves the model by assuming that the distribution of tastes for the population and the distribution of adjustment costs for land follow a Fréchet distribution.

[9] To minimize notation, we assume no distortions in the economy. One could easily add frictions to our model, as to generate a gap in the wage of workers between agricultural and manufacturing sector, for example. The main results and intuitions of our model in that case would remain the same.

[10] The endogenous productivity growth could come from different mechanisms, such as knowledge accumulation,

Appendix A1 describes how we solve for the equilibrium in the model. In equilibrium, output prices in ℓ^o equalize the ones in the rest of the country, as represented by location ℓ^*. The share of workers choosing to live in ℓ^o depends on how real wages in ℓ^o evolve relative to the rest of the country ℓ^*. Lastly, real wages in ℓ^* increases over time, as the rest of the country experiences continuous manufacturing productivity growth.

2.2 Rural Locations

If ℓ^o is a rural location, it has a comparative advantage in agriculture relative to the rest of the country in the initial period—i.e., $A_{A,0}\left(\ell^o\right)/A_{M,0}\left(\ell^o\right) > A_{A,0}\left(\ell^*\right)/A_{M,0}\left(\ell^*\right)$. In that case, in $t = 0$, location ℓ^o specializes in agriculture, employing a larger share of workers and land in agriculture relative to ℓ^*. The evolution of the manufacturing sector in a rural location is therefore disadvantaged relative to ℓ^*: Because ℓ^o has a smaller share of manufacturing workers relative to ℓ^*, manufacturing productivity and real wages grow faster in ℓ^* intead of ℓ^o, inducing workers to gradually move away from the rural location ℓ^o.

Consider now a positive, *permanent* agricultural productivity shock in ℓ^o in period $t' > 0$. As a result, real wages increase in ℓ^o at time t', which stanches its outflow of workers. Meanwhile, this shock reduces even further the share of workers in manufacturing, slowing down manufacturing productivity growth in ℓ^o, which harms the growth of local real wages in the long-run. Since there is a small share of manufacturing workers to begin with, this latter mechanism tends to have a small effect on real wages, so that in the long-run the rural location still has a larger population when its hit by the shock, relative to a scenario in which there is no shock.

innovation, and local capital investments in buildings and structures. The formulation that we adopt here, which is based on a production externality, gives us a simple way of capturing these mechanisms. Alternatively, one could also assume that the local productivity growth comes from a combination of local investments and positive externalities related to productivity growth in the rest of the economy, as in Desmet et al. (2018).

2.3 Urban Locations

If ℓ^o is an urban location, it has a comparative advantage in manufacturing relative to the rest of the country in the initial period—i.e., $A_{A,0}(\ell^o)/A_{M,0}(\ell^o) < A_{A,0}(\ell^*)/A_{A,M}(\ell^*)$. An urban location ℓ^o tends to have a larger proportion of workers in manufacturing relative to ℓ^*, which generates faster growth of manufacturing productivity and real wages relative to the rest of the country ℓ^*. As a result, in contrast to a rural location, an urban location absorbs workers over time.[11]

Here, a positive, *permanent* agricultural productivity shock in period t' also increases the share of agricultural workers, which leads to a reduction in the share of manufacturing workers, and consequently a reduction in the speed of manufacturing productivity growth. However, relative to a rural location, this latter mechanism is dominant, because the share of manufacturing workers is large. As such, in the long-run, the urban location absorbs fewer workers from elsewhere when it is hit by the shock, and the population is smaller in ℓ^o relative to a scenario in which there is no shock.[12]

2.4 A Region with Multiple Locations

Lastly, we study a region I, which contains a set of small, open economy locations, some of which are rural, and one of which is urban. We think of region I as the area of a town plus its hinterland—or, alternatively, we think of a regional market with several towns and villages. We assume that workers draw an additional taste shock for locations within region I, so that they choose between locations within region I, and between region I and the rest of the country ℓ^*.[13]

The evolution of the economy in I depends on the total number of rural locations: If the

[11]That occurs even if the urban location starts in disadvantage relative to ℓ^* in terms of its real wages: Since this location has faster manufacturing productivity growth, over time, workers still move away from ℓ^* and into ℓ^o.

[12]If the agricultural productivity shock is large enough, the urban location becomes a rural one, so that that $A_{A,t'}(\ell) > A_{A,t}(\ell^*)$, and it starts releasing workers to the rest of the economy ℓ^*.

[13]Specifically, we assume that taste shocks are drawn from a nested Fréchet structure (see Farrokhi and Pellegrina (2022) for an application of this type of distribution), so that we allow larger elasticity of substitution between locations within region I relative to the outside option of living in the rest of the economy ℓ^*. See appendix for details.

number of rural locations is small, then region I as a whole has a faster productivity growth than the rest of the country, which leads to an inflow of workers to I. If we have a larger number of rural locations instead, then the opposite occurs. Additionally, these migration patterns depend on how much workers prefer to live in regions from I, relative to moving to other parts of the country.

The impact of a positive, *permanent* agricultural productivity shock in a region depends on the proportion of rural locations in region I. For example, if the region only contains rural locations, it will pull workers from the rest of the country. If instead there is an urban location in the region, the shock will affect how urban and rural locations interact within that region (via migration linkages), and the aggregate impact on the population of the region will depend on whether the absorption of workers in villages will dominate the release of workers from the urban location. Later, we consider this type of spatial unit in our empirical analysis.

To conclude this section, we underscore that our model indicates that a *positive*, permanent agricultural productivity shock increases the share of agricultural workers *both* in urban and rural locations.[14] Next, we relate these types of locations to towns and villages, respectively, in the context of India. We find empirically that in both types of locations, in line with our model's prediction, the share of agricultural workers rises with irrigation—which is consistent with empirical findings in Foster and Rosenzweig (2004). Our model, however, indicates that there are starkly different effects of the agricultural shock on population, depending on whether the shock hits an urban or a rural location, or whether the shock hits a broader region with multiple types of locations. Indeed, we find empirical evidence consistent with this prediction of the model.[15]

[14] Simulations of the stylized model indicate that quantitatively the increase in the share of agricultural workers should be small in villages, since the initial share of agricultural workers is already large, but potentially large for towns. Indeed, when we go to data, we find results that are consistent with these predictions.

[15] If we had a service sector oriented towards local consumers in the model, there would occur an increase in its size as a result of the agricultural productivity shock in villages, since real wages would rise in the local economy. In towns, however, we would observe a drop in the size of the service sector, since real wages would drop in the long-run.

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.

3.1 Demographic and Economic Censuses

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.

3.2 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.[16] 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 error (Donaldson and Storeygard, 2016).

3.3 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).[17] Attribute data on canals is completed using Central Water Commissions' Management Information System of Water Resources Projects and India WRIS Wiki.[18] 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).

3.4 Summary Statistics

Appendix Table 1 reports the sample size and descriptive statistics. Our analysis encompasses approximately 1,500 irrigation projects (i.e., command areas), for which we have high-resolution

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

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

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

data on the boundaries and all other relevant geographic features. The sample 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.

4 Empirical Strategy

Our empirical strategy exploits the discontinuity in program inclusion arising at the boundary of command areas, comparing villages (and 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 the 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 the lowest elevation lines on both sides of the ridge and the terminus of the main canals. Using elevation data, we confirm that the command area boundaries are essentially flat, with average slopes on the order of a 20 cm decline per 100 meters distance.

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 km from the boundary (see Figure 1), but our results are not affected by the choice of a narrower or wider bandwidth.[19] In addition,

[19] 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.

we restrict the sample to locations (villages and towns) that were less than 30 sq km, in order to remove the largest cities which are unlikely to be affected by irrigation, though our results are not sensitive to relaxing this restriction.

Formally, our main estimation takes the form:

$$y_{i,d,p,b} = \alpha + \beta C_i + \mathbf{X_i}\mathit{\Gamma} + \nu_d + \mu_{p,b} + \varepsilon_{p,i}, \tag{3}$$

Agricultural and non-agricultural outcomes are denoted by y_{ipdb}, where i is an index for location (village or town) located within 10 km of irrigation project p in district d and b is an index for the nearest 5- or 10-km boundary segment of the project's command area.

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. We also control for a vector of village geographic characteristics, X_i, which includes altitude, ruggedness, distance to a 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. All regressions include district fixed effects ν_d. In addition, they include fixed effects $\mu_{p,b}$, that in different estimations designate either the specific command area, or the 5-km or 10-km segment of the command area boundary to which the location i is nearest. This helps ensure the only relatively proximate treated and control localities are compared.

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, for villages, 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

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).

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 km. Fourth, we winsorize the outcome 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. (2022) and Blakeslee et al. (2019), who evaluate specific surface irrigation projects in Rwanda and India, respectively.

4.1 Units of Observation: Villages, Towns, and Geographic Cells

The model presented in the preceding section motivates a separate analysis for villages, towns, and aggregated geographic cells that may contain zero or more of both types of location. We therefore estimate treatment effects for villages and towns separately.

Appendix Table 1 present summary statistics for towns and villages. In column (1) are given means characteristics in villages, and in columns (2)–(4) the differences between towns and villages, with column (3) including command area fixed effects, and column (4) additionally restricting the sample to locations smaller than 30 sq km. Villages and towns have starkly different economic and demographic characteristics: towns have much larger populations, smaller agricultural sectors, more large firms (per capita), and greater household asset holdings. In light of our model, we would expect the impact of the treatment to differ between these two types of locations.

In addition, we estimate impacts at a unit of observation that aggregates outcomes over geographic cells that may contain several villages and towns. We work with two aggregations. First, we take the unit of observation to include a single town and all the villages within 10 km of its boundary (its "hinterland"). Because towns located in the command area will generally have larger shares of their hinterland in the command area as well (see Appendix Figure A3.2), this will capture the aggregate impact on a town and its affiliated rural areas of an increase in agricultural productivity.[20] Outcomes of interest (e.g., population, number of employees in firms) are aggregated, and geographic controls averaged, over all locations (villages and towns) located within this cell. Treatment status, as before, is based on the location of the town centroid; its hinterland, however, may include place on either side of the command area boundary.

Second, we construct the cells to include the area located either up to 10 km inside, or 10 km outside, of every 10-km segment of the command area boundaries. There are therefore two roughly 10×10-km cells per boundary segment, one of which is "treated" by the irrigation project and one of which is not. Here too, outcomes and control variables are aggregated (or averaged, depending on the variable) over all locations (villages and towns) located within the cell.

4.2 Town Formation

One potential concern with the separation of our analysis for villages and towns is that whether a village graduates to the status of being a town could be affected by the irrigation schemes, which may confound the estimation of treatment effects with composition effects. This concern is made more salient by the fact that approximately half the towns in 2011 only became towns after 1991 (Appendix Figure A3.3), which was after the vast majority of irrigation schemes had been completed.

[20] Appendix Figure A3.2 plots the share of the 10- and 20-km circles surrounding a town that is in the command area against the distance of the town (centroid) from the command area boundary. The shares using the 10-km circles are 21% and 65% for control and treatment towns, respectively; and 26% and 53% when using the 20-km threshold.

To assess the likelihood of the treatment causing villages to become towns, it is necessary to understand how these entities are defined. Formally, a location is designated as a town if it satisfies three criteria: (1) a population greater than 5,000; (2) a population density greater than 400 per sq km; and (3) a male agricultural labor share less than 25%. As shown in Table 1, and as discussed in greater detail in Appendix A2, there are stark differences between villages and towns across a variety of economic and demographic characteristics. Importantly, these differences remain even when comparing the smallest towns with the largest villages (columns (4)-(6) of Appendix Table A3.1 and Appendix Figure A3.4).

Appendix Table A3.2 presents several tests for endogenous town formation. In column (1), we restrict the sample to villages and towns that were close to meeting the criteria for being classified as towns, and estimate the impact of being in the program areas on attaining township status.[21] 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 1921 and 1951, respectively. In columns (4)–(8), we use as the unit of observation the 10×10-km cells on either side of the command area boundary, as described in section 4.1. In columns (4)-(6) we take as the outcome the number of towns per 100 square km in 1921, 1951, and 2011, respectively; and in columns (7) and (8) the change between 1921–2011 and 1951–2011.

We find no evidence for differential town formation across the control and treatment groups. Essentially, this means that the 6.1% increase in village population in treatment areas, which we report below—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 starkly different demographic and economic characteristics of villages and towns. Though there is no evidence for endogenous town formation, in Section 5.6 we report additional tests that account for any violations of this finding, by restricting the analysis to locations that are unlikely to have been formed due to irrigation. The results of this analysis are very similar to the benchmark results.

[21] These are villages and towns that had: (1) a population between 4,000 and 6,000 people; a population density of more than 350 persons per sq km; and male agricultural labor force that is less than 30%.

4.3 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 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), and for which there is not an adjacent river (within 500 meters), no such trend breaks are visible (blue lines in Figure 2). We therefore use this sampling restriction in our analysis.

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 A3.3).[22]

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

[22]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.

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.

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

5.1 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 report results for villages and towns in parallel, and show that the effects on agricultural activity are substantial and similar for both types of locations.

We illustrate the results for villages graphically in Figure 3.1–3.3, which plots these outcomes against the indicated distance bins from the boundary using specification 3, labelling distance as negative within the command area and positive outside of it, and excluding villages that overlap the command area boundary. Results for regressions without controls are depicted in Appendix Figure A3.6.1–A3.6.3. All three outcomes display clear discontinuities at the boundary.

We report regression estimates for various agricultural outcomes in Tables 2 and 3. 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.).[23] 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 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 3).

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.

5.2 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.

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

Similar to the illustration for agriculture outcomes, we present our results for urbanization and development in villages graphically in Figure 3.4–3.6. Results for regressions without controls are depicted in Appendix Figure A3.6.4–A3.6.6. All three outcomes display clear discontinuities at the boundary.

Figure 4.1 and Table 4 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, 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 bench-marking them against the modest (13%) effect on irrigated area, implying irrigation elasticities for these outcomes of substantial magnitudes.[24]

5.3 Labor Force Composition

In Figure 4.2 and Table 5, we document the impact of canal irrigation on labor force participation and composition using demographic census data. We find a small positive effect on agricultural labor in villages, though it loses significance with the inclusion of boundary-segment fixed effects. In towns, we estimate a substantial 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.

[24] To provide a better sense of magnitudes, we can use a simple migration equation to infer the implicit changes in wages, which we think as a *proxy* for labor productivity. Specifically, consider the following equation for the supply of workers in a region:

$$\left(\frac{N_1}{N_0}\right)^{1/\theta} = \frac{w_1}{w_0}$$

where N_1 and N_0 are the population with and without the treatment, respectively, w_1 and w_0 are the wages, and θ is the migration elasticity. Using a migration elasticity of 2, which is common in the literature, and the results from Table 4, which indicate that treated villages and towns experience population changes of +7% and -18%, respectively, we obtain wage changes of +3% for villages and -9% for towns. Bustos et al. (2016), for comparison, find that a 1 standard deviation in the productivity of soybeans increases labor productivity by 13%.

5.4 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 6 in greater detail. Employment in firms increases by 5.8% in villages, with effects evident for manufacturing (4.6%) and service firms (7.2%). These effects seem to be driven by small firms (less than 10 workers), which we associate with smaller shops serving local consumers. 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.

5.5 Assets

Figure 4.4 and Table 7 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 no evidence for corresponding effects on asset holdings in towns.

5.6 Additional Discussion of Identification and Robustness

We perform several additional estimations that provide indirect tests of our empirical approach. First, Appendix Table A3.4 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 A3.5 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 A3.6, Panel A presents estimates for villages while Panel B reports estimates for towns. A comparison of these results with those from equation 3 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 A3.7 we also use alternative bandwidths ranging from 2–30 km. 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 outcomes 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 A3.7).

Sixth, we estimate our main results while removing villages which intersect the boundaries (see Appendix Table A3.8, Panel A); with winsorized outcome variables at the 5th and 95th percentiles (Appendix Table A3.8, 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 A3.8, 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 Figures 3.5 and 3.6).

In addition, though we find no evidence for endogenous town formation, we nonetheless conduct robustness test for the town analysis to assess whether any such endogeneity could have biased our results. First, we restrict the town sample to those locations (villages and towns) that met the township criteria given by the census (population, population density, and male agricultural labor force). Second, we restrict the sample to locations (villages and towns) that continue to meet the township criteria after reducing their populations by 6.1% (the estimated increase in village population in treatment areas), which removes locations that may have crossed the village-town population threshold due to the treatment. Third, we restrict the sample to locations (villages and towns) that were well above the census criteria, in order to exclude all locations that could have possibly converted from villages to towns due to the intervention. The results, given in Appendix Table A3.9, are largely unchanged from the baseline analysis.

5.7 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 A3.1, 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).

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 from towns, but in their vicinity become negative: villages within 2 km 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 8, 9, and 10 present corresponding estimates using a (treatment-interacted) binary indicator for town-proximity which takes a value of 1 for villages within 4 km of the nearest town. Table 11 reports results for household assets and home amenities.

5.8 Aggregate Geographic Units

The heterogeneity in impacts across towns and villages raises the important question of what the local aggregate impacts are of the agricultural productivity shocks. The model suggests that the impacts on total population will depend on the ratio of towns to villages and the degree of frictions to labor mobility.

To conduct the aggregate analysis, we use the two approaches discussed in Section 4.1 based on local geographic cells: the first composed of a town and its surrounding 10-km hinterland; and the second using 10×10-km cells on either side of the command area boundary. The regression is specified similarly to equation 3, with the index i now designating a whole cell. For the 10×10-km cell analysis, we also include a binary indicator for the presence of a town in the cell, and its interaction with the treatment term.

The results of these analyses are given in Table 12. The results indicate that, on aggregate, the command area experiences increases in population density, firm employment, and manufacturing employment; but no change in employment in large firms. However, there are substantial declines for all these outcomes when there is a town present, indicating that losses occurring in towns are not offset by gains to surrounding villages.

5.9 Discussion

It important to emphasize that our estimates, whether at the village, town or cell level, capture the *local* economic impacts of agricultural productivity gains offered by irrigation. The widespread introduction of irrigation also has general equilibrium country-wide impacts, including the potential acceleration or slowdown in aggregate structural transformation, but these do not lend themselves to causal inference using our approach. The local impacts we estimate occur against that backdrop and in addition to it. For example, if general equilibrium impacts have driven aggregate declines (increases) in manufacturing and large-firm production at the country level, they have done so more in the treatment (control) area of the study sample than in the control (treatment).

Though our RDD empirical strategy limits the conclusions one can draw about the mechanisms driving the observed impacts on population, our model provides guidance about this question. The model predicts that, overall, relative increases in manufacturing productivity will move workers from villages to towns. A permanent increase in agricultural productivity in a given village will slow the outward movement of workers from that village, causing a *relative* (if not absolute) increase in its population, in comparison to unaffected villages. The same shock in a town will reduce the inward movement of workers, causing a relative decline in the town population in comparison to unaffected towns.

In principle, the observed increase in village population could also arise because of an increase in in-migration or in the native population growth rate. We find reduced out-migration, as posited by our model, to be the most plausible explanation, given the predominance of rural-urban migration in male migration patterns in India,[25] and the scale of rural-urban migration occurring during these years. To put our estimated 6.1% village population increase in perspective, using a back-of-the-envelope calculation based on changing urbanization rates in India between 1971–2011 (and excluding urbanization caused by village conversion to towns), we estimate that the average Indian village would have been approximately 11% larger in 2011 without rural-to-urban migration.[26] For the same reason, we find reduced in-migration to be the most plausible explanation of the decline in town population, rather than increased out-migration or reduced population growth.

Our model makes no assumptions about the spatial extent of migration. Workers can move to irrigated towns from both irrigated and non-irrigated villages within the study sample, as well

[25]Using census migration data from 2011, we estimate that more than 75% of male work-related migration is to urban areas.

[26]Numbers used in this calculation can be found in Bhagat (2018).

as from more distant locations outside the study sample; and the same is true for non-irrigated towns. The spatial extent at which migration occurs in practice will depend on frictions in labor mobility. The fact that we find a negative population effect for aggregate (10-km) geographic cells that contain towns suggests that these migration flows are likely not confined to nearby locations.[27]

Asher et al. (2021) seek to empirically estimate general equilibrium population impacts on the India-wide spatial distribution of towns and urban populations. Using a difference-in-differences strategy—comparing town growth in the command area and adjacent areas to that occurring farther away from the command area—the authors find evidence for greater urbanization in the vicinity of the command areas. The control group in this analysis is primarily composed of towns outside our study sample, while our own control group is sufficiently close to the command area that most of the towns are considered to be treated in the difference-in-differences analysis.[28] We focus instead on local impacts that are more amenable to causal inference using the spatial break in project coverage, but view the findings of these two approaches as complementary.

6 Conclusion

Over much of the 20th century, the construction of large-scale surface irrigation infrastructure was one of the most capital-intensive investments undertaken 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

[27]In addition, the lack of any village population trend with distance outside the command area (see Figure 3.4) suggests that there has not been migration from control- to treatment-area villages, as the incentive to migrate would vary with distance from the boundary and likely create a trend in village population.

[28]The measure of treatment status for the Asher et al. (2021) analysis is the share of the 20 km area surrounding a town which is located in a command area, which is used both as a continuous variable as well as an indicator taking a value of 1 for towns for which the value is above 20%. Within our study sample, the mean value of the surrounding-area share variable is 26% for towns *outside* the command area (i.e., our control group), with 59% of these towns being above the 20% threshold (see Appendix Figure A3.2). An analysis using the full set of towns (in-sample and out-of-sample) would therefore include in its control group primarily towns well outside the command area, and in its treatment group both within-sample (control and treatment) towns, as well as towns deep inside the command area which were excluded from our analysis. It is important to reiterate that we find no evidence for differential town formation across the control and treatment areas of our study sample.

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. In rural areas, irrigation increases population density, night light density, and built-up area, while also modestly increasing per-capita wealth.

In towns (and villages close to towns), however, population density, nightlight density, and the non-agricultural labor force share are reduced in irrigated areas; and there is a decline in employment in firms, including manufacturing and large firms. When aggregating outcomes across a broader area including both towns and villages, we find that the presence of a town causes population and economic losses that are not offset by gains in nearby villages. 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 the shock.

The ability to simultaneously conduct our analysis at a fine 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. Due to the local nature of the treatment effects being identified by the spatial RDD, we are unable to capture the economy-wide impacts of irrigation expansion on structural transformation, economic growth, and the spatial allocation of labor. We therefore interpret our findings as reflecting the local, long-term effects of irrigation occurring against the benchmark of these economy-wide impacts; and, similarly to Foster and Rosenzweig (2004) and Bustos et al. (2016), avoid making claims about latter.

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.

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.

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.*

Bhagat, R. B. (2018): *Urbanization in India: Trend, pattern and policy issues*, International Institute for Population Sciences.

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.

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

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.

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.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018): "The geography of development," *Journal of Political Economy*, 126, 903–983.

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.

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

Farrokhi, F. and H. S. Pellegrina (2022): "Trade, technology, and agricultural productivity," *Working Paper.*

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. M. Wingender (2021): "Two blades of grass: The impact of the green revolution," *Journal of Political Economy*, 129, 2344–2384.

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.

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 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., F. Kondylis, J. Loeser, and J. Magruder (2022): "Factor market failures and the adoption of irrigation in rwanda," *American Economic Review*, 112, 2316–52.

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.

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.

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

Mukhopadhyay, P. (2017): "Does administrative status matter for small towns in India?" in *Subaltern urbanisation in India*, Springer, 443–469.

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.

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.

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.

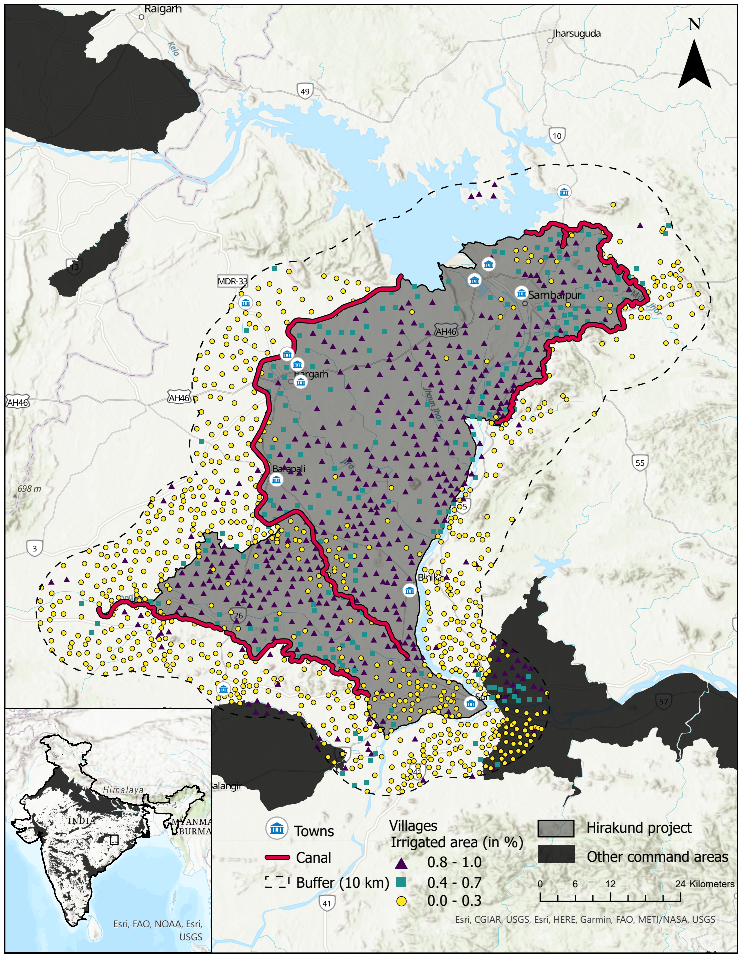
Studwell, J. (2013): *How Asia works: Success and failure in the world's most dynamic region*, Open Road+ Grove/Atlantic.

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

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.

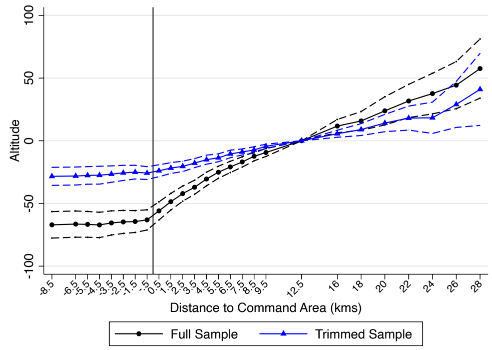
Figures

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

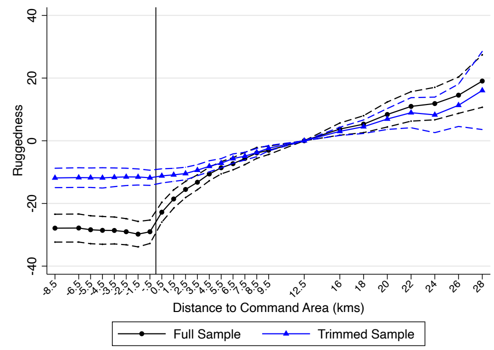


Notes: The empirical strategy compares villages on either side of the command area border (shaded light grey) in a 10-km buffer (denoted by the dotted black line). To compare nearby villages, 5-km boundary segment fixed effects are used, which are calculated by splitting the border into smaller parts. (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.

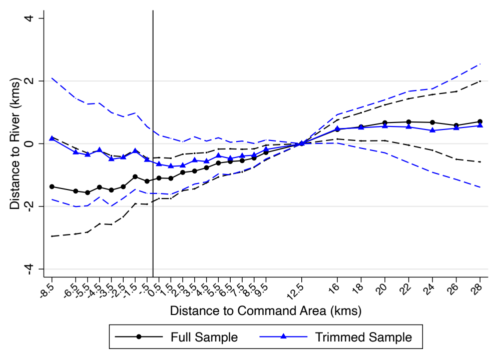
Figure 2: Geographic Features



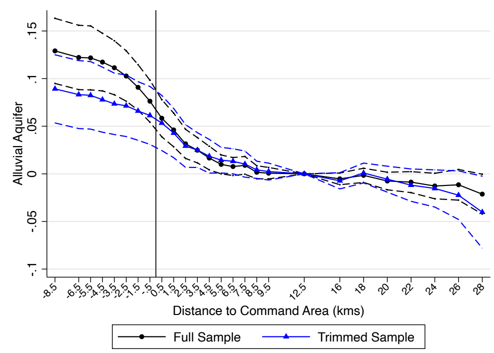
2.1: Altitude



2.2: Ruggedness



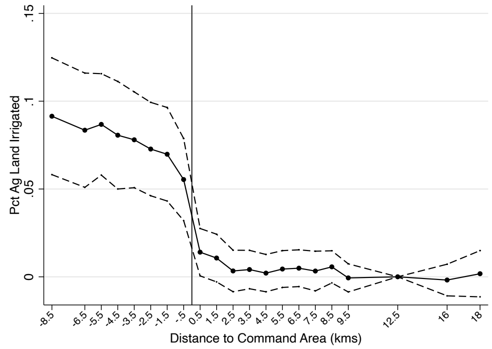
2.3: Distance to River



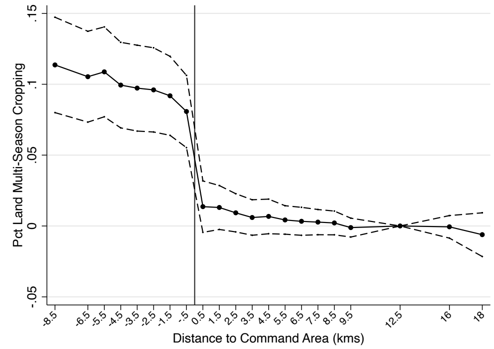
2.4: Aquifer

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 km), and Figure 2.4 depicts whether a village lies on top of an alluvium/water-deposited aquifer.

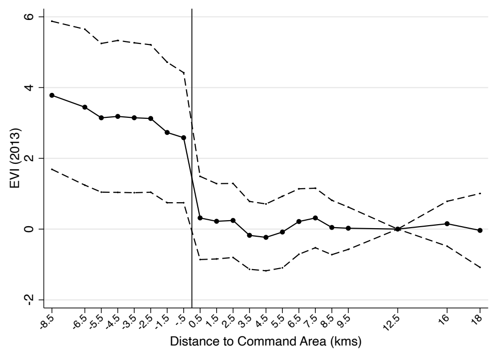
Figure 3: Agriculture and Development



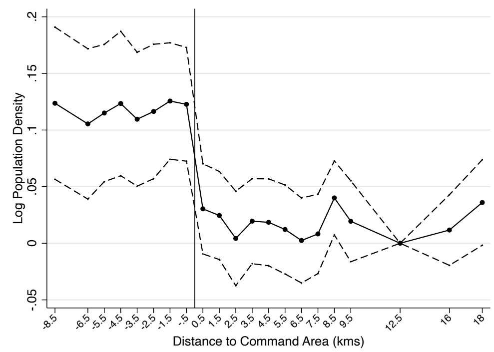
3.1: Pct of Agriculture Area Irrigated



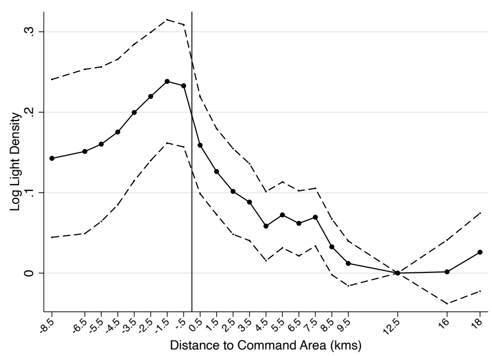
3.2: Multi-Season Cropping



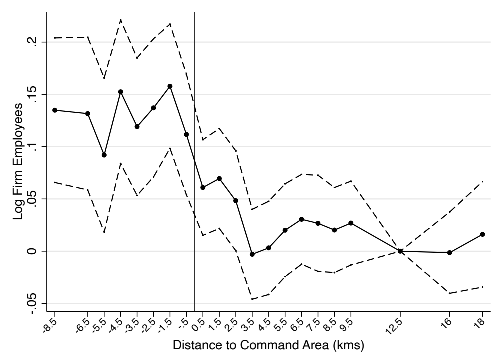
3.3: Dry Season Vegetation



3.4: Log Population Density



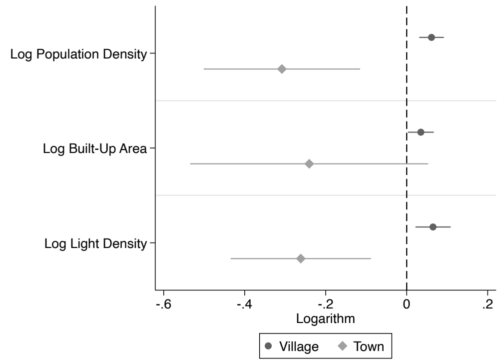
3.5: Log Light Density



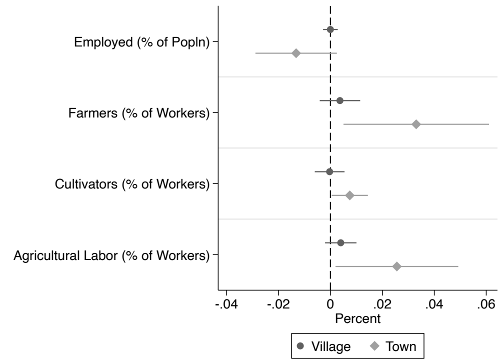
3.6: Log Firm Employment

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.

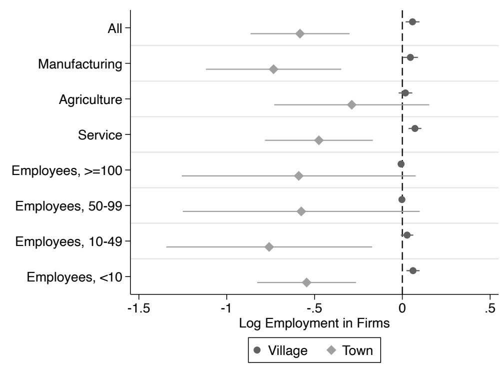
Figure 4: Labor Force Participation, Firm Activity and Assets



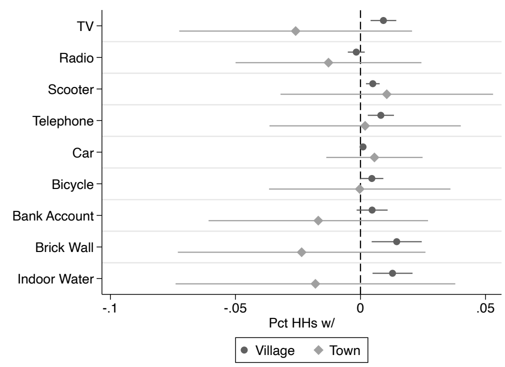
4.1: Urbanization



4.2: Labor Force Participation



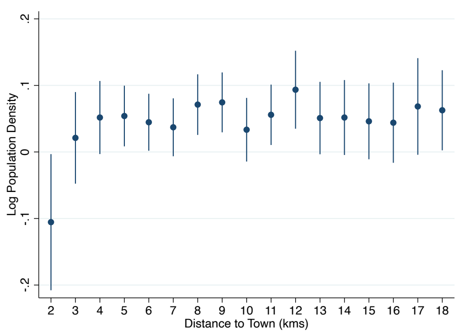
4.3: Employment in Firms



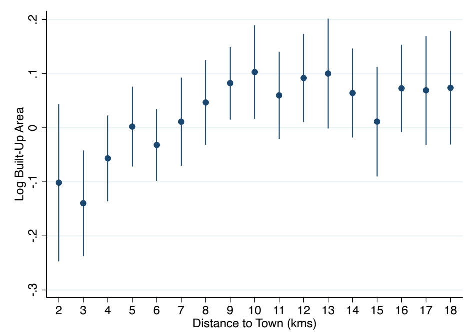
4.4: Assets

Notes: This figure plots β from equation 3 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).

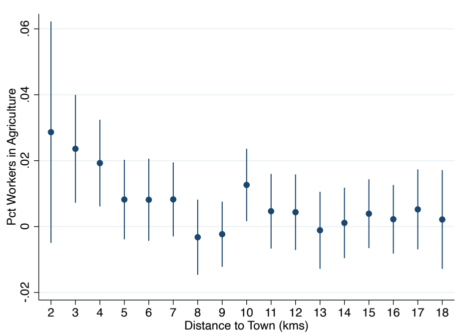
Figure 5: Treatment Effect by Distance to Town



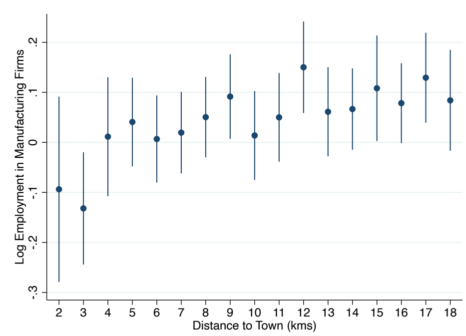
5.1: Log Population Density



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: 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 sq km				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 2: 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: $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \mathit{\Gamma} + \nu_d + \eta_p + \varepsilon_i$ (column 1) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \mathit{\Gamma} + \nu_d + \mu_b + \varepsilon_i$ (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 3: 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 4: 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 5: 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 6: 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 6 – 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 7: 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 7 – 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 8: Urbanization in Villages (by Proximity to Town)

	Log		
	Population Density (1)	Built-up Area (2)	Light Density (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 km 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 9: 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 km 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 10: 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 km 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 11: Assets and Housing, Villages (by town proximity)

	Assets							Housing			
	Pct						Pct	Pct			
	TV	Radio	Scooter	Phone	Car	Bicycle	Bank Account	Brick Wall	Inside Water	Condition Good	Num 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 km 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.

Table 12: Area Impacts

	Level of Analysis			
	Town + Nearby Villages	Project Boundary 10 x 10 km cells		
	(1)	(2)	(3)	(4)
Panel A: Log Population Density				
Treatment	-0.110**	0.100***	0.089***	0.107***
	(0.038)	(0.014)	(0.013)	(0.013)
Town Present			0.913***	1.016***
			(0.035)	(0.046)
Treatment × Town Present				-0.205***
				(0.050)
Control Mean	6.787		5.436	
R-squared	0.763	0.662	0.740	0.741
N	1467	11702	11702	11702
Panel B: Log Employment				
Treatment	-0.151**	0.108***	0.093***	0.109***
	(0.067)	(0.021)	(0.019)	(0.019)
Town Present			1.362***	1.458***
			(0.043)	(0.060)
Treatment × Town Present				-0.191***
				(0.069)
Control Mean	8.680		6.253	
R-squared	0.696	0.599	0.689	0.689
N	1467	11648	11648	11648
Panel C: Log Employment Firms >50 Workers				
Treatment	-0.690**	0.012	0.000	0.023
	(0.327)	(0.019)	(0.018)	(0.017)
Town Present			1.045***	1.177***
			(0.053)	(0.078)
Treatment × Town Present				-0.262***
				(0.094)
Control Mean	4.609		0.869	
R-squared	0.497	0.364	0.438	0.440
N	1467	11648	11648	11648
Panel D: Log Employment Firms Manufacturing				
Treatment	-0.212**	0.084***	0.068***	0.085***
	(0.100)	(0.020)	(0.018)	(0.018)
Town Present			1.408***	1.508***
			(0.045)	(0.062)
Treatment × Town Present				-0.198***
				(0.074)

Continued on next page

Table 12 – Continued from previous page

	(1)	(2)	(3)	(4)
Control Mean	7.261		4.112	
R-squared	0.691	0.507	0.614	0.614
N	1467	11648	11648	11648

Notes: Table reports results from two estimating equations: $y_{ipdb} = \alpha + \beta C_i + \delta TownPresent_i + \kappa(C_i \times TownPresent_i) + \mathbf{X}_i \Gamma + \nu_d + \eta_p + \varepsilon_i$ (columns 1) and $y_{ipdb} = \alpha + \beta C_i + \mathbf{X}_i \Gamma + \nu_d + \mu_b + \varepsilon_i$ (columns 2, 3 and 4) where, y_{ipdb} is an aggregated outcome of interest in a town and its surrounding 10-km hinterland i (column 1) or in a 10 $\times$ 10-km cell i in irrigation project p in district d along boundary segment b (columns 2, 3 and 4); C_i is an indicator variable for whether the centroid of the town/cell 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 10-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 outcome in Panel A is derived from the Census of India 2011. It reports ln(population density). The outcomes in Panels B, C and D are derived from Economic Census 2012-13: Panel B report ln(total employment) in all enterprises/firms. Total employment = agriculture + manufacturing + services. Panel C report reports ln(employment) in firms with more than 50 workers; and panel D reports ln(manufacturing sector employment). 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: Model

This section characterizes the equilibrium of the model presented in the main body of the paper. We start by presenting workers and landowner choices, given the distributional assumptions and adjustments costs. We then present the equations characterizing the equilibrium prices in location ℓ^*, which is a location representing the rest of the country. Next, we show the equilibrium in the small, open economy location ℓ^o and in the case in which we have a region with several locations. To save on notation, we drop index of region and time in what follows, unless otherwise indicated.

A1.1 Landowners and workers choices

In the case in which we have a single location ℓ^o and the rest of the country ℓ^*, we index locations by $\ell \in \{\ell^o, \ell^*\}$. We assume that the distribution of tastes in the population from location ℓ^o is distributed according to a Fréchet distribution with dispersion parameter κ. In that case, optimal location choices for workers are given by:

$$N\left(\ell^o, \ell\right) = \frac{w\left(\ell\right)^{\kappa}}{w\left(\ell^o\right)^{\kappa} + w\left(\ell^*\right)^{\kappa}} N,$$

where $N\left(\ell^o, \ell\right)$ is the total mass of workers from ℓ^o living in ℓ, $w\left(\ell\right)$ is the wage, and N the total population of workers who choose between ℓ^o and ℓ^*.

For landowners, we also assume that distribution of the adjustment costs follow a Fréchet distribution with dispersion parameter θ. This gives the following mass of land allocated to sector k:

$$L_k\left(\ell\right) = \frac{r_k\left(\ell\right)^{\theta}}{r_A\left(\ell\right)^{\theta} + r_M\left(\ell\right)^{\theta}} L\left(\ell\right).$$

To simplify the solution to the model, we assume that all adjustment costs are paid back to landowners.

For the third type of geographic unit, in which a region is composed of several urban and rural locations. Here, we index locations using ℓ_i^o, in which i denotes a sub-location within region I. We assume that tastes are drawn from a nested Fréchet distribution (see Farrokhi and Pellegrina, 2022). Specifically, workers draw their taste parameters based on a hierarchical structure, in which the taste for living in locations ℓ_i^o within I agains location ℓ^* comes from an upper tier, and the taste shock for living in locations

i within I comes from a lower tier. As such, the share of workers living in ℓ_i^o is:

$$N\left(\ell^o, \ell_i^o\right) = \frac{w\left(\ell_i^o\right)^{\tilde{\kappa}}}{\tilde{w}^{\tilde{\kappa}}} \cdot \frac{\tilde{w}^{\tilde{\kappa}}}{\tilde{w}^{\kappa} + w\left(\ell^*\right)^{\kappa}} N$$

where $\tilde{w} = \left(\sum_{i \in I} w\left(\ell_i^o\right)^{\tilde{\kappa}}\right)^{\frac{1}{\tilde{\kappa}}}$ and the share of workers living in ℓ^* is

$$N\left(\ell^o, \ell^*\right) = \frac{w\left(\ell^*\right)^{\kappa}}{\tilde{w}^{\kappa} + w\left(\ell^*\right)^{\kappa}} N.$$

A1.2 Equilibrium in the rest of the country

To simplify the solution of the equilibrium in ℓ^*, we assume that the population N is small relative to the entire population in the rest of the country, N^*—as such, the location choices of N do not affect wages and land rents in ℓ^*.[29] With this simplifying assumptions, we can solve for the equilibrium in the rest of the country ℓ^* in any period independently of the location choices of N.

For the rest of the country ℓ^*, marginal productivity of workers must equalize between sectors, which gives

$$\frac{p_A\left(\ell^*\right)}{p_M\left(\ell^*\right)} = \frac{A_M\left(\ell^*\right)}{A_A\left(\ell^*\right)} \frac{\left(1-\alpha_M\right)}{\left(1-\alpha_A\right)} \frac{L_M\left(\ell^*\right)^{\alpha_M}}{L_A\left(\ell^*\right)^{\alpha_A}} \frac{N_A\left(\ell^*\right)^{\alpha_A}}{N_M\left(\ell^*\right)^{\alpha_M}} \tag{A1}$$

Labor market clearing gives

$$N_k\left(\ell^*\right) = \mu_k N^* \tag{A2}$$

Firms' FOC give

$$\frac{N_k\left(\ell^*\right)}{L_k\left(\ell^*\right)} = \frac{1-\alpha_k}{\alpha_k} \frac{r_k\left(\ell^*\right)}{w\left(\ell^*\right)} \tag{A3}$$

Landowners optimal choice give

$$\frac{L_{k'}\left(\ell^*\right)}{L_k\left(\ell^*\right)} = \left(\frac{r_{k'}\left(\ell^*\right)}{r_k\left(\ell^*\right)}\right)^{\theta} \tag{A4}$$

[29] One can easily extend the model to incorporate several locations, each having some influence on the price of every other location. Here, we make these assumptions with the goal of simplifying the solution of the model and make intuition clearer.

Combining equations (A2), (A3) and (A4), we get

$$\frac{L_M(\ell^*)}{L_A(\ell^*)} = \left[\frac{\mu_M \alpha_M}{1-\alpha_M} \Big/ \frac{\mu_A \alpha_A}{1-\alpha_A}\right]^{\frac{\theta-1}{\theta}} \tag{A5}$$

Combining equations (A2) and (A5) with full labor and land employment, we get the equilibrium values $\tilde{L}_M(\ell^*)$, $\tilde{L}_A(\ell^*)$, $\tilde{N}_A(\ell^*)$, and $\tilde{N}_M(\ell^*)$. With these equilibrium values, we normalize the price of manufacturing to 1 and recover the price of agriculture:

$$p_A(\ell^*) = \frac{A_M(\ell^*)}{A_A(\ell^*)} \frac{(1-\alpha_M)}{(1-\alpha_A)} \frac{\left(\tilde{L}_M(\ell^*)/\tilde{N}_M(\ell^*)\right)^{\alpha_M}}{\left(\tilde{L}_A(\ell^*)/\tilde{N}_A(\ell^*)\right)^{\alpha_A}}. \tag{A6}$$

Lastly, we recover equilibrium wages based on:

$$w(\ell^*) = p_M(\ell^*)^{1-\mu_M} A_M(\ell^*)(1-\alpha_M)\left[\frac{\tilde{L}_M(\ell^*)}{\tilde{N}_M(\ell^*)}\right]^{\alpha_M}. \tag{A7}$$

The equations above provide the equilibrium values for the economy at any given period t. The evolution of manufacturing productivity, $A_{M,t}(\ell^*)$, follows equation (2).

A1.3 Equilibrium in a location

In a small, open economy location ℓ^o, output prices are fixed at the country level. The equations that we use to solve for the model in each period t are as follows. First, from marginal productivity of land and labor, we get

$$w(\ell^o) = p_k^* A_k(\ell^o)(1-\alpha_k)\left(\frac{L_k(\ell^o)}{N_k(\ell^o)}\right)^{\alpha_k} \tag{A8}$$

$$r_k(\ell^o) = p_k^* A_k(\ell^o)\,\alpha_k\left(\frac{N_k(\ell^o)}{L_k(\ell^o)}\right)^{1-\alpha_k} \tag{A9}$$

Second, from firms' FOC we get

$$N_k(\ell^o) = \frac{\frac{1-\alpha_k}{\alpha_k} r_k(\ell^o) L_k(\ell^o)}{\frac{1-\alpha_A}{\alpha_A} r_A(\ell^o) L_A(\ell^o) + \frac{1-\alpha_M}{\alpha_M} r_M(\ell^o) L_M(\ell^o)} N(\ell^o, \ell) \tag{A10}$$

Third, landowners optimal choices give

$$L_k\left(\ell^o\right)=\frac{r_k\left(\ell^o\right)^\theta}{r_A\left(\ell^o\right)^\theta+r_M\left(\ell^o\right)^\theta}L\left(\ell^o\right) \tag{A11}$$

Fourth, optimal location choices give

$$N\left(\ell^o,\ell\right)=\frac{\left(w\left(\ell\right)\right)^\kappa}{\left(w\left(\ell^o\right)\right)^\kappa+\left(u\left(\ell^*\right)\right)^\kappa}N. \tag{A12}$$

Using the five equations (A7) to (A12), we solve for five endogenous variables, $w\left(\ell^o\right)$, $r_k\left(\ell^o\right)$, $N_k\left(\ell^o\right)$, $L_k\left(\ell^o\right)$, and $N\left(\ell^o,\ell\right)$ in each period t. The evolution of manufacturing productivity, $A_{M,t}\left(\ell^o\right)$, then follows equation (2).

A1.4 Equilibrium in a region with many types of locations

As in the previous case, prices are fixed at the country level. The equations that we use to solve for the model are now as follows. First, marginal productivity of labor gives

$$w\left(\ell_i^o\right)=p_k^*A_k\left(\ell_i^o\right)\left(1-\alpha_k\right)\left(\frac{L_k\left(\ell_i^o\right)}{N_k\left(\ell_i^o\right)}\right)^{\alpha_k}$$

$$r_k\left(\ell_i^o\right)=p_k^*A_k\left(\ell_i^o\right)\alpha_k\left(\frac{N_k\left(\ell_i^o\right)}{L_k\left(\ell_i^o\right)}\right)^{1-\alpha_k}$$

Second, from firms' FOC we get

$$N_k\left(\ell_i^o\right)=\frac{\frac{1-\alpha_k}{\alpha_k}r_k\left(\ell_i^o\right)L_k\left(\ell_i^o\right)}{\frac{1-\alpha_A}{\alpha_A}r_A\left(\ell\right)L_A\left(\ell_i^o\right)+\frac{1-\alpha_M}{\alpha_M}r_M\left(\ell_i^o\right)L_M\left(\ell_i^o\right)}N\left(\ell^o,\ell_i^o\right)$$

Third, landowners optimal choices give

$$L_k\left(\ell_i^o\right)=\frac{r_k\left(\ell_i^o\right)^\theta}{r_A\left(\ell_i^o\right)^\theta+r_M\left(\ell_i^o\right)^\theta}L_k\left(\ell_i^o\right)$$

Fourth, optimal location choices give

$$N\left(\ell^o, \ell_i^o\right) = \frac{w\left(\ell_i^o\right)^{\tilde{\kappa}}}{\tilde{w}^{\tilde{\kappa}}} \cdot \frac{\tilde{w}^{\tilde{\kappa}}}{\tilde{w}^{\kappa} + w\left(\ell^*\right)^{\kappa}} N,$$

where

$$\tilde{w} = \left(\sum_{i \in I} w\left(\ell_i^o\right)^{\tilde{\kappa}}\right)^{\frac{1}{\tilde{\kappa}}}.$$

We solve for $w\left(\ell_i^o\right)$, $\tilde{w}$, $r_k\left(\ell_i^o\right)$, $N\left(\ell_i^o\right)$, $N_k\left(\ell_i^o\right)$, $L_k\left(\ell_i^o\right)$, and $N\left(\ell^o, \ell_i^o\right)$ in each period t. The evolution of manufacturing productivity, $A_{M,t}\left(\ell_i^o\right)$, follows equation (2).

A1.5 Simulations

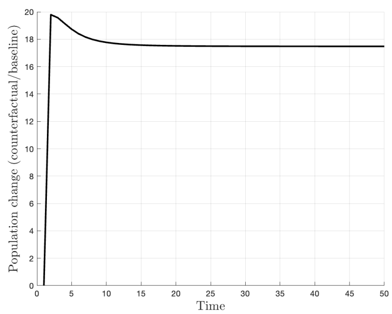
Appendix Figure A.1 shows the impact of a positive, permanent agricultural productivity shock in the three types of units of analysis discussed in Section 2. Specifically, starting the economy from $t = 1$, we increase permanently agricultural productivity in period $t = 2$ by 10%. Each figure presents the impact of the shock on population relative to a counterfactual scenario in which there is no shock in that location. We refer to the counterfactual scenario as the economy hit by the shock, and the baseline scenario as the economy without the shock.

Panel (a) reports the impact of the agricultural productivity shock on a rural location. In that case, there is an increase the inflow of workers right in after the shock, relative to the baseline scenario. As a consequence, however, there is a reduction in the share of workers in the manufacturing sector, so that manufacturing productivity growth slows down in the counterfactual location. As such, over time, this initial advantage of the counterfactual economy hit by the shock, relative to the baseline one, is attenuated. Since the manufacturing sector is small to begin with, this initial advantage is never lost over time due to its productivity growth: The rural economy under the shock remains with a larger population in the long-run.

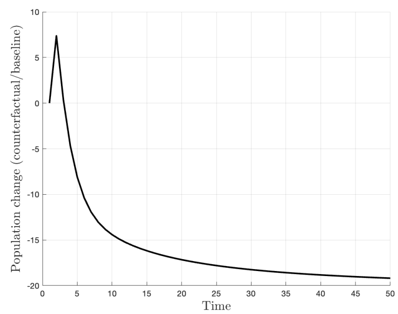
Panel (b) turns to the impact on a urban location. Here, an increase in agricultural productivity provides an initial advantage to the urban location hit by the shock—notice that, right when the location is hit by the shock, the agricultural productivity is larger and the manufacturing productivity is the same compared to the baseline economy. Because of this increase in agricultural productivity, this location absorbs more workers relative to the baseline scenario with no shock in the short-run. However, as the economy unfolds, because productivity growth is slower in the counterfactual, the initial advantage given by the productivity shock is lost and the region becomes smaller relative to the baseline scenario without the shock.

Panel (c) presents the impact on a region composed of multiple urban and rural locations. Here, the region as a whole is increasing its population and manufacturing employment over time. As such, the impact of the agricultural productivity shock on productivity growth dominates the overall impact on the region's population. Naturally, if the region is fully composed of rural locations, we would have the opposite result. The extent to which the impact of an agricultural productivity shock on a region will reduce or increase population depends on the composition of the locations.

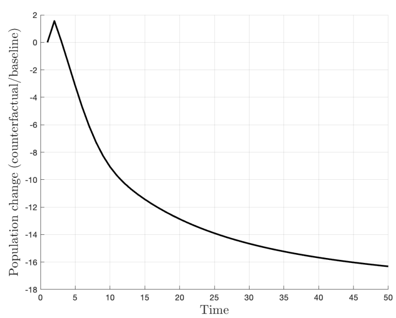
Figure A.1: Simulations of the Impact of an Agricultural Productivity Shock



A.2.1: Rural Location



A.2.2: Urban Location



A.2.3: Region

Notes: This figure presents simulations of the stylized model with and without and a permanent increase in agricultural productivity of 10% in period $t = 2$. The y-axis in each figure shows the percentual change in population between the simulation with the shock and the one without it, in each period in timer. The x-axis shows the time. Panel (a) documents the impact on a rural location, Panel (b) on an urban location, and Panel (c) on a region with multiple rural and urban locations.

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).[30] 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.[31] 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.

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

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

[31] 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.

various agricultural activities, the strength of the economic census lies in providing firm-level information on employment for non-agricultural establishment.[32] 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 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).

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

Towns and Villages. An important element of our analysis is the estimation of impacts separately for towns and villages. It is therefore important to clearly articulate the administrative, demographic, and economic characteristics that distinguish towns from villages, and the process through which villages may turn into towns over time.

Formally, Indian settlements are classified as villages or towns, and towns are further classified as either "statutory towns" or "census towns." A statutory town is a formally incorporated township, which is an administrative unit defined by law and governed by an urban local body.[33] A census town is a category used for census enumeration. Census towns continue to be administered as villages, but are classified as towns due to their higher population, population density, and non-agricultural labor share. Formally, the Census of India classifies a location as a census town if the following three conditions are met: the population exceeds 5,000; the population density is more than 400 persons per sq km; and less than 25% of the main male working population is employed in the agricultural sector (alternatively, more than 75% of the male working population should be employed in the non-agricultural sector).[34]

Appendix Table 1 depicts some of the key differences between villages and towns (statutory and census types combined). In column (1) 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. The table highlights the starkly different character of towns and villages. Towns are larger and more populous. They also have much larger shares of built-up area and smaller shares of agricultural area, higher light density and population density, and greater asset holdings and household amenities. Towns also feature lower agricultural labor force shares, and greater numbers of employees in formal service and manufacturing firms, of which a larger share is also employed in large firms with more than 10 or 100 workers.

Appendix Table A3.1 reports parallel statistics that are now differentiated by the two types of towns. As is apparent, though census towns are substantially smaller than statutory towns, they are otherwise

[33] Local governance institutions in statutory towns are empowered to collect property taxes to provide their citizens with civic services such as piped water, waste management, drainage, etc. They may also develop rules to regulate land use and land development.

[34] Among towns, 73% meet these criteria; while, amongst villages, 0.30% do so. In robustness tests, we find that the results for towns are unchanged when restricting the sample to all locations (towns and villages) possessing these characteristics.

remarkably similar,[35] and both differ markedly from villages along virtually every dimension. In columns (4)-(6), we restrict the sample to villages and towns with populations of 4,000-6,000. Importantly, even where their populations are similar, towns differ dramatically from villages.[36]

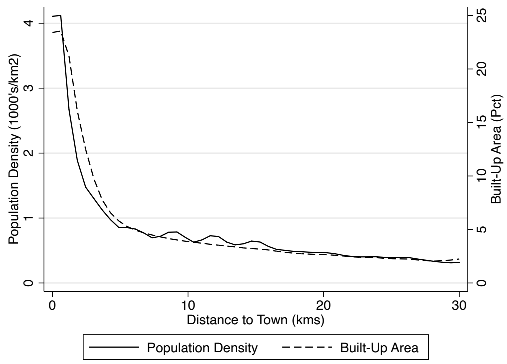
As highlighted in Table A3.1, towns are the locus for most large-scale production, particularly manufacturing. In addition, towns play an important role as “market centers” to surrounding villages, being nodes for long-distance trade and the site for the production and procurement of locally produced non-tradables. Appendix Figure A3.1 depicts the spatial structure of rural economies, with village demographic and economic characteristics (population density, the composition of the labor force, and employment in formal firms) plotted against the distance to the nearest town. While differences between villages is small in comparison to those between villages and towns, the plot clearly shows that villages located are closer to towns are less agricultural, more urbanized and economically developed.

[35]In a similar vein, Mukhopadhyay (2017) also argues that there is little difference between census towns and statutory towns.

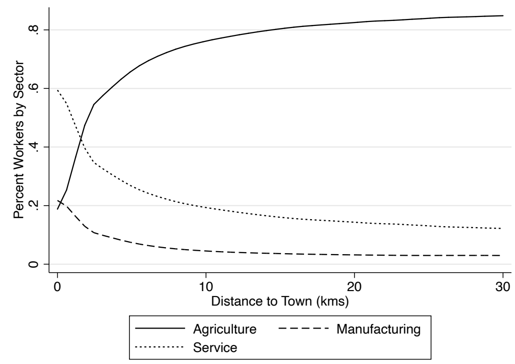
[36]Differences are smaller, of course, for area, population, and the number of employees in different firm categories because of the sample population restriction.

A3 Appendix: Additional Figures and Tables

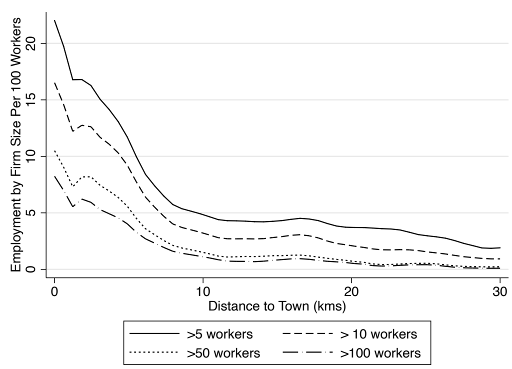
Figure A3.1: Spatial Distribution of Economic Activity



A3.1.4: Urbanization



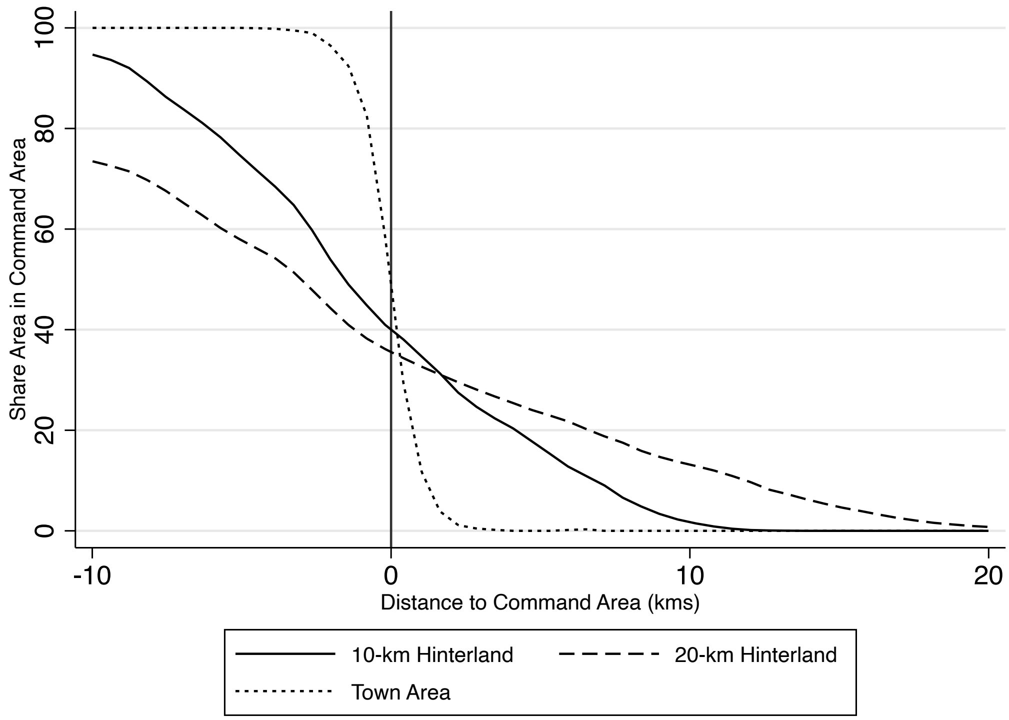
A3.1.5: Labor force



A3.1.6: Firm Size

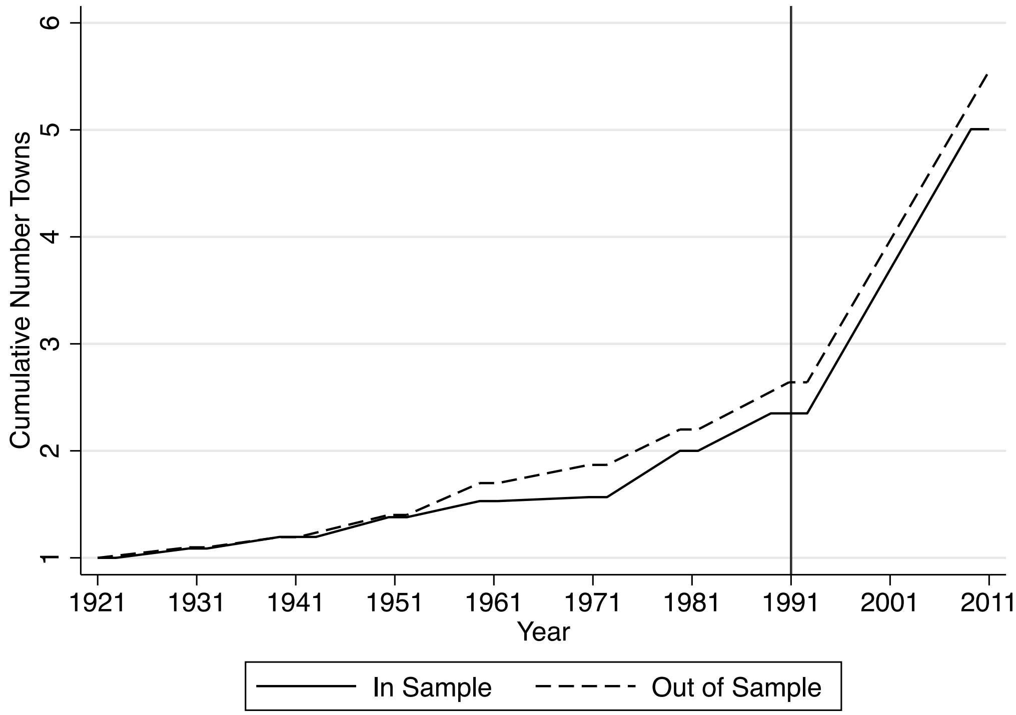
Notes: The above figure plots the spatial distribution of economic activity of villages relative to towns. Distance from village centroid to the nearest town (in km) is on the x-axis. Figure A3.1.4 depicts the population density per 1,000 square km (Census of India 2011) on the left y-axis and percentage of built-up area on the right y-axis. Figure A3.1.5 depicts percent of workers in agriculture (Census of India 2011), manufacturing and service sectors (Economic Census 2012-13). Figure A3.1.6 depicts employment by firm size (Economic Census 2012-13).

Figure A3.2: Command Area Exposure



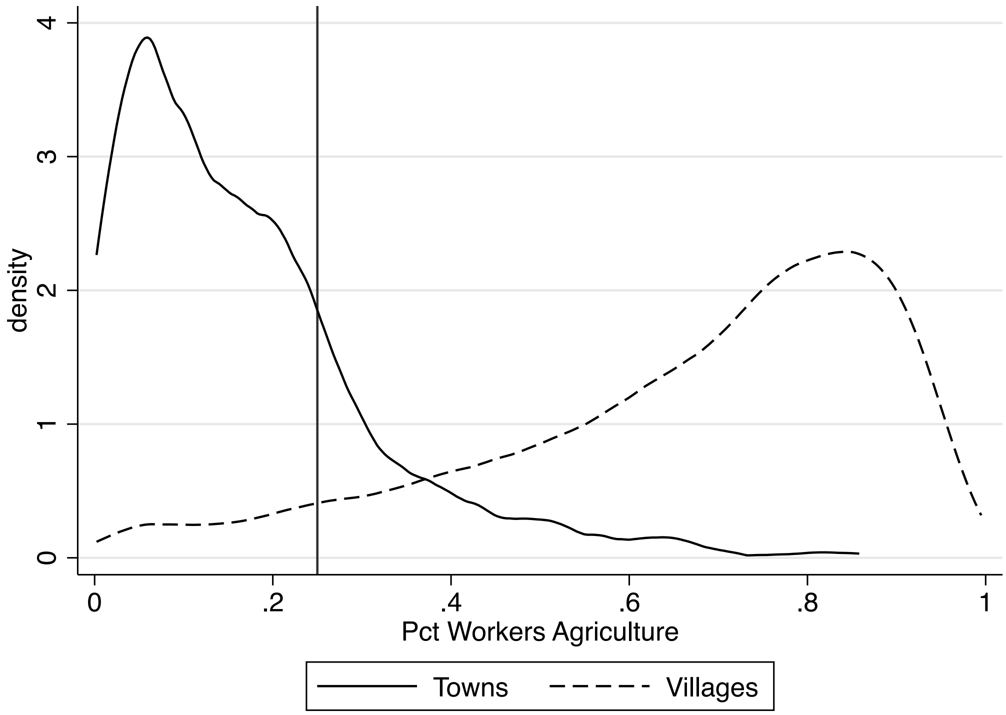
Notes: This figure plots the share of the town area, and the area surrounding the town ("hinterland"), that is within the command area, against the distance of the town centroid to the command area boundary. Hinterland share is shown for areas extending 10 and 20 km from the town boundaries, respectively.

Figure A3.3: Town Presence



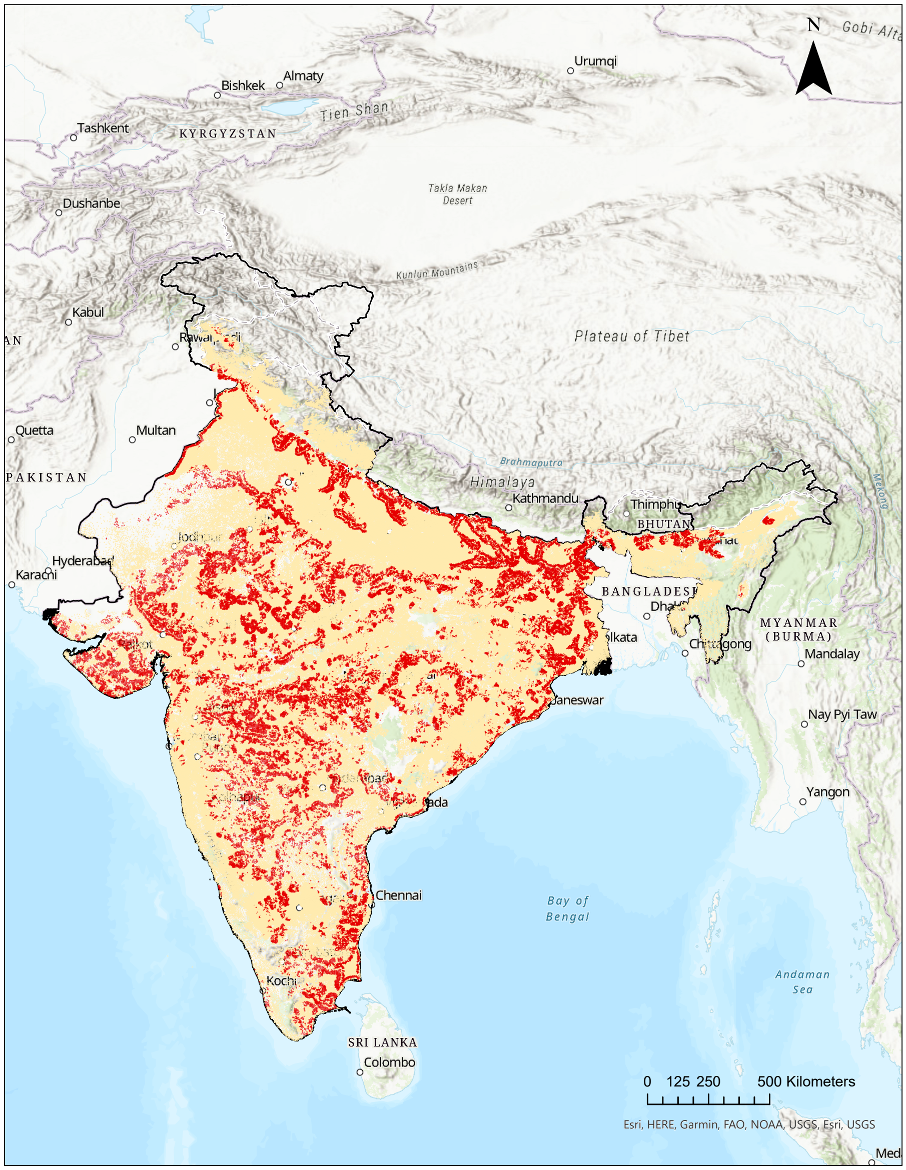
Notes: This figure plots the number of towns (demeaned by the number in 1921) against year. The sample is disaggregated into “in sample” towns, which are those located in the control and treatment groups from the study sample; and “out of sample” towns, which are towns located outside of the study sample (excluding out-of-sample towns located within command areas).

Figure A3.4: Distribution of Agricultural Workforce, Matched Town/Village Sample



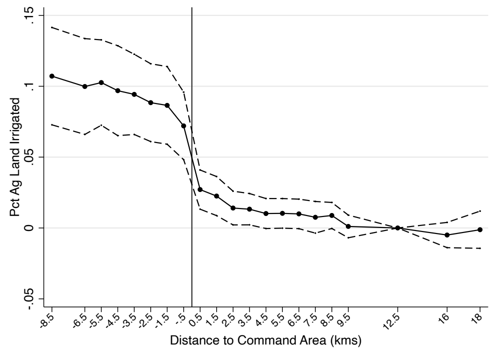
Notes: This figure gives the distribution of share of workers engaged in agriculture for towns and villages. The sample is restricted to towns and villages with a population of 4,000-6,000 individuals.

Figure A3.5: Villages in the trimmed sample

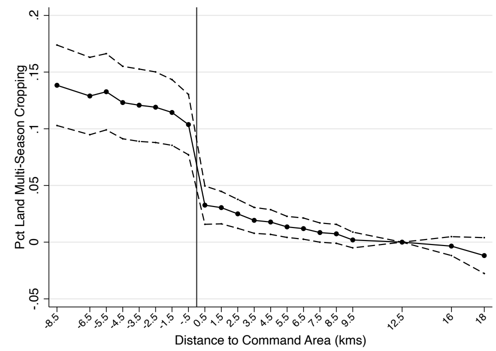


Notes: Each dot corresponds to a village in India. The villages in the study sample are denoted by a red dot, while those not in the study sample are denoted in topaz sand color. There are approx 145,000 villages in the trimmed sample, which account for 22 percent of the nearly 650,000 villages in the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

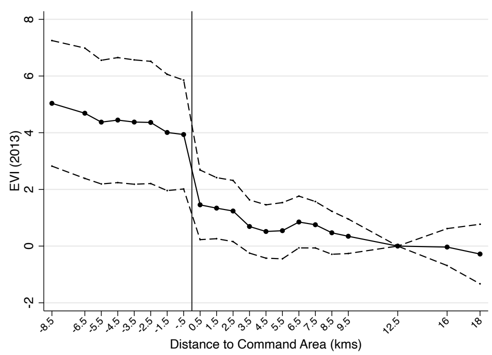
Figure A3.6: Agriculture and Development, w/o Geographic Controls



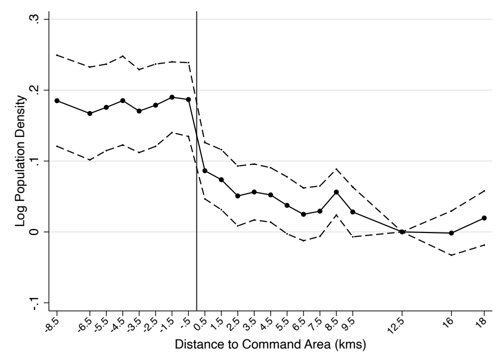
A3.6.1: Pct of Agriculture Area Irrigated



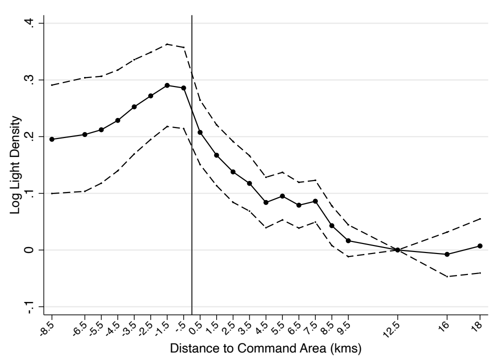
A3.6.2: Multi-Season Cropping



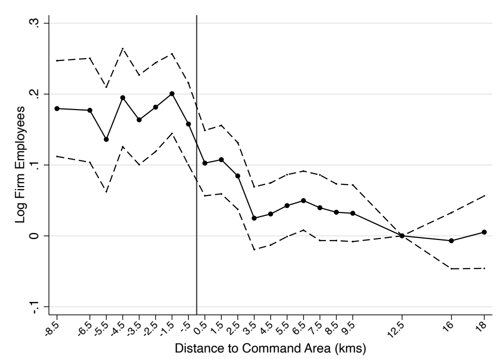
A3.6.3: Dry Season Vegetation



A3.6.4: Log Population Density



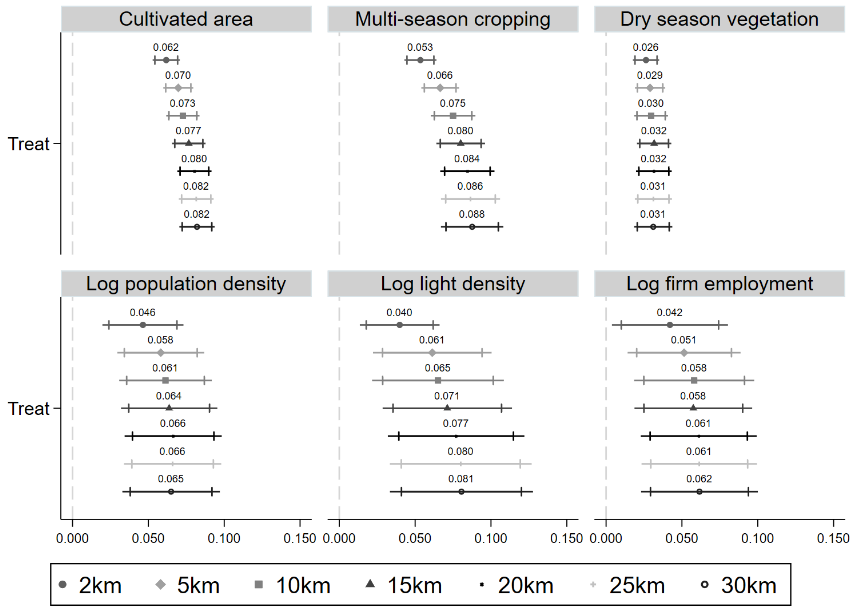
A3.6.5: Log Light Density



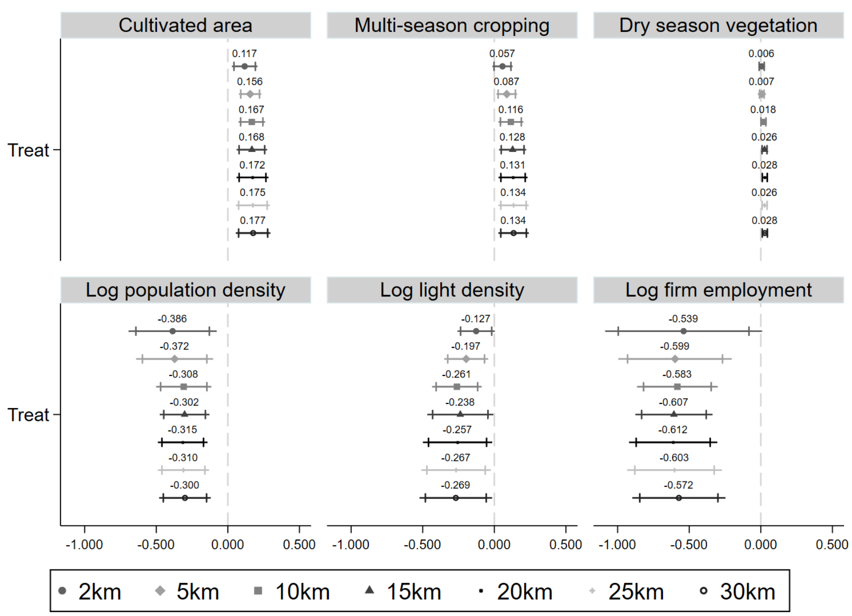
A3.6.6: Log Firm Employment

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 A3.6.1 depicts area under irrigation as percent of cultivable land; Figure A3.6.2 depicts land area that is cropped twice or thrice as percentage of agricultural area; and Figure A3.6.3 depicts dry season vegetation indices as percentage of total village area. Figure A3.6.4 depicts log population density; Figure A3.6.5 depicts mean nighttime lights per sq km; Figure A3.6.6 depicts number of employees in firms across manufacturing, agriculture and services enterprises.

Figure A3.7: Robustness to varying bandwidths



A3.7.1: Villages



A3.7.2: 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.

Table A3.1: Summary Statistics by Town Type

	Village Mean	Town – Village: Census Towns	Town – Village: Formal Towns	Matched Size Sample: Village Mean	Matched Size Sample: Town – Village	
	(1)	(2)	(3)	(4)	(5)	(6)
Total Area (km2)	4.093	2.559**	13.544***	9.918	4.361	3.529
		(1.023)	(0.963)		(5.788)	(5.659)
Share Area Built-Up	0.050	0.162***	0.196***	0.067	0.107***	0.106***
		(0.012)	(0.008)		(0.012)	(0.012)
Share Area Agriculture	0.625	-0.199***	-0.154***	0.616	-0.177***	-0.177***
		(0.032)	(0.011)		(0.030)	(0.031)
Light Density	6.075	18.514***	14.534***	7.865	11.354***	11.245***
		(1.262)	(0.645)		(1.063)	(1.081)
Tot Population (1,000s)	1.618	10.545***	66.653***	4.809	0.474***	0.000
		(0.917)	(5.005)		(0.041)	(0.000)
Population Density (1,000s/km2)	0.700	2.679***	4.131***	1.094	1.344***	1.344***
		(0.227)	(0.212)		(0.166)	(0.169)
Pct Male Workers Ag	0.757	-0.515***	-0.507***	0.687	-0.372***	-0.371***
		(0.013)	(0.012)		(0.025)	(0.025)
Employees in Firms (100s)	1.365	21.947***	105.651***	4.218	3.665***	3.317***
		(6.294)	(7.901)		(1.091)	(1.094)
Employees in Manu Firms (100s)	0.291	9.749***	25.791***	0.924	2.396***	2.300**
		(3.273)	(2.301)		(0.927)	(0.921)
Share Employees in Firms >10 Workers	0.060	0.070***	0.094***	0.097	0.024	0.024
		(0.012)	(0.006)		(0.017)	(0.018)
Share Employees in Firms >100 Workers	0.007	0.033***	0.039***	0.017	0.024*	0.025*
		(0.006)	(0.004)		(0.013)	(0.013)
Pct HHs w/TV	0.282	0.213***	0.220***	0.337	0.161***	0.162***
		(0.014)	(0.012)		(0.022)	(0.022)
Pct HHs w/Telephone	0.522	0.163***	0.149***	0.571	0.123***	0.122***
		(0.010)	(0.007)		(0.016)	(0.016)
Pct HHs w/Scooter	0.143	0.090***	0.086***	0.143	0.055***	0.056***
		(0.010)	(0.005)		(0.010)	(0.010)
Pct HHs w/Brick Wall	0.473	0.236***	0.173***	0.556	0.172***	0.173***
		(0.025)	(0.009)		(0.020)	(0.020)
Pct HHs w/Water Source on Premises	0.217	0.371***	0.405***	0.292	0.202***	0.199***
		(0.039)	(0.017)		(0.050)	(0.050)
Population Control						Yes

Note: This table reports descriptive statistics for villages and towns. Column (1) reports the mean for villages; column (2) reports the mean difference between census towns and villages; and column (3) reports the mean differences between municipalities and villages. Columns (4)-(6) restrict the sample to villages and towns with 4,000-6,000 inhabitants. Column (4) reports the mean for villages in this sample; column (5) reports the mean difference between towns and villages in this sample; and column (6) reports the mean difference while include a control for total population.

Table A3.2: Town Formation

	Town-Level Analysis			Project-Level Analysis (per $100km^2$)				
	Town	Town in		Num Towns			Δ 2011–	
	Status	1921	1951	1921	1951	2001	1921	1951
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.015	-0.002	0.023	-0.004	0.002	-0.023	-0.019	-0.025
	(0.055)	(0.028)	(0.032)	(0.010)	(0.012)	(0.028)	(0.025)	(0.024)
Control Mean	0.307	0.160	0.213	0.053	0.072	0.296	0.243	0.225
R-squared	0.353	0.433	0.464	0.184	0.184	0.342	0.354	0.362
N	411	1513	1513	12791	12791	12791	12791	12791
Town Sample	Yes	Yes	Yes					
Village Sample	Yes							
Marginal Sample	Yes							
Area Sample				Yes	Yes	Yes	Yes	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 sq km. In columns (2) and (3), the sample is limited to all towns and the outcome variable is an indicator for having been a township in 1921 and 1951, respectively. Columns (4)–(6) take as the outcome the number of towns per 100 sq km in 1921, 1951, and 2011; and in column (7) and (8) the change between 1921–2011 and 1951–2011. This analysis is conducted at the 10×10-km cell level. Fixed effects are included for the project and district. Geographic controls are included taken as the mean values for the cells. 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 A3.3: Balance, Geographic Features

	Control Mean	Difference: Full Sample	Difference: Trimmed 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 km), 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 A3.4: Treatment Effects at Canal Boundaries

	Pct				Log			Pct HHs/w	
	Ag Area Canal	Ag Area Irrigated	Area 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 A3.5: Placebo Regressions, 1991

	Pct Ag Land Irrigated		Log		Pct	Pct Worker	Log
	Canal	Any	Pop Density	Light Density	Workers	Ag	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 A3.6: Robustness tests using boundary distance controls

	Pct				Log			Pct HHs/w	
	Ag Area Canal	Ag Area Irrigated	Area 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			0.130***	-0.048	-0.430*	-0.353***	-0.782***	-0.039	-0.087**
	NA		(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 A3.7: Treatment Effects excluding Villages <2 km Inside Boundary

	Pct				Log			Pct HHs/w	
	Ag Area Canal	Ag Area Irrigated	Area 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 A3.8: Robustness Tests

	Pct				Log			Pct HHs/w	
	Ag Area Canal	Ag Area Irrigated	Area 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.

Table A3.9: Town Impacts, Alternative Samples

	Log Population Density	Log Employees by Firm Type			Pct Workers Agriculture
		All	Manu	>50 Work	
	(1)	(2)	(3)	(4)	(5)
Panel A: Met Criteria					
Treatment	-0.174***	-0.243**	-0.280*	-0.601**	0.037***
	(0.067)	(0.114)	(0.153)	(0.236)	(0.009)
R-squared	0.522	0.528	0.543	0.390	0.422
N	1353	1342	1342	1342	1353
Panel B: Met Criteria (Adjusted Population)					
Treatment	-0.174**	-0.251**	-0.611**	-0.287*	0.037***
	(0.069)	(0.114)	(0.238)	(0.157)	(0.009)
R-squared	0.528	0.524	0.392	0.537	0.413
N	1284	1273	1273	1273	1284
Panel C: Non-Marginal Towns					
Treatment	-0.184**	-0.246**	-0.293**	-0.586**	0.028***
	(0.071)	(0.105)	(0.136)	(0.255)	(0.007)
R-squared	0.562	0.538	0.562	0.413	0.460
N	872	867	867	867	872

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 to all locations (villages and towns) that met the township criteria of (1) more than 5000 inhabitants, (2) a population density above 400 inhabitants per sq km, and (3) a male labor force share less than 25% agricultural. Panel B restricts the sample to all locations (villages and towns) that met the township criteria after reducing their populations by 6.6% (the estimated village population impact). Panel C restricts the sample to all locations (villages and towns) that were well above the township criteria, specifically: (1) more than 7000 inhabitants, (2) a population density above 500 inhabitants per sq km, and (3) a male labor force share less than 20% agricultural. Column 1 ln(population density) from Census of India 2011; Columns 2–4 ln(total employed workers) in all enterprises, manufacturing enterprises, and enterprises with more than 50 workers, respectively, from Economic Census 2012-13; and Column 5 percent of all workers in agriculture. 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.