STEG WORKING PAPER

THE LONG-RUN DEVELOPMENT IMPACTS OF AGRICULTURAL PRODUCTIVITY GAINS: EVIDENCE FROM IRRIGATION CANALS IN INDIA

Sam Asher, Alison Campion,
Douglas Gollin, and Paul Novosad

JUNE 2022
STEG WP004

The Long-Run Development Impacts of Agricultural Productivity Gains: Evidence from Irrigation Canals in India*

Sam Asher† Alison Campion‡ Douglas Gollin§ Paul Novosad¶

June 27, 2022

Abstract

How do investments in agricultural productivity translate into development and structural transformation? We estimate the long-run impacts of India's irrigation canals, which span 300,000+ km and deliver water to 130,000+ villages. Drawing on high-resolution data on every household, firm, village, and town in India, we use three empirical strategies to characterize the direct and spillover effects of large increases in agricultural productivity. Our findings are consistent with a spatial equilibrium model in which labor is mobile, and urban areas have non-farm productivity advantages. In the long run, areas directly treated by canal irrigation have sharply higher agricultural productivity and population density, but similar non-farm employment shares to non-canal areas. Persistent consumption gains accrue only to landowners and structural transformation occurs almost exclusively through the concentrated growth of regional towns. In the long run, the substantial productivity effects of canals were equilibrated through the movement of labor across space rather than within locations across sectors.
JEL Codes: O13, O18, Q15

*We are grateful for excellent research assistance from Sam Besse, Aditi Bhowmick, Kritarth Jha, Toby Lunt, Shraddha Mandi, and Sankalp Sharma. Gayatri Acharya, Richard Damania, Anju Gaur, Ijsbrand de Jong, Stuti Sharma, Vivek Srivastava and Esha Zaveri at the World Bank provided invaluable insight into the water and agriculture sectors in India, as well as assistance in obtaining data. We appreciate the feedback from seminar and workshop participants at UC Berkeley, Stanford, the University of Chicago, IIM Ahmedabad, IFPRI, BREAD, Dartmouth College, University of Exeter, University of Geneva, the World Bank, STEG, NEUDC, UC Davis, and UC Merced. Financial assistance from the World Bank Agriculture Global Practice and Research Support Budget, as well as Emergent Ventures (Mercatus Center, George Mason University), is gratefully acknowledged.

†Imperial College London, sam.asher@imperial.ac.uk

‡Development Data Lab, campion@devdatalab.org

§University of Oxford, douglas.gollin@qeh.ox.ac.uk

¶Dartmouth College, paul.novosad@dartmouth.edu

1 Introduction

The link between agricultural productivity and structural transformation has long been a central concern of development economics (Nurkse, 1952; Lewis, 1954; Ranis and Fei, 1961; Schultz, 1964). Early authors such as Johnston and Mellor (1961) and Jorgenson (1961)—echoed later by Mellor (1986), Timmer (1988), and World Bank (2007)—argued that agricultural productivity growth was an essential precursor for broader structural transformation and long-run economic growth. This early literature held that productivity growth in agriculture could have the seemingly paradoxical effect of shrinking the agricultural sector as a share of the total economy. Building on the insight that food is an essential good for the poor, agricultural development economists generated a class of models in which countries that are unproductive in agriculture must devote large shares of labor and other resources to meet their food needs.[1] However, the literature offers both theoretical and empirical challenges to this view of the relationship between agricultural productivity and sectoral change. For example, Matsuyama (1992) showed that in an open economy, increases in agricultural productivity can cause *specialization* in agriculture via comparative advantage, while Bustos et al. (2016) show that the direction of local structural transformation can depend on the factor bias of the technical change in agriculture.

This paper studies one of the most significant episodes of agricultural productivity change of the past two centuries: the construction of India's massive irrigation canal network. This network of canals—artificial channels that carry water into dryland areas for application to crops, primarily during the dry winter growing season—spans over 300,000 km and serves over 130,000 villages, nearly one in four in India. Canals were historically the most important source of irrigation in India, and even in the 21$^{\text{st}}$ century they are the second largest source of irrigation in India after groundwater, providing water to agricultural regions with over a quarter billion inhabitants. In 2011, fully 57% of rural Indians lived within 10 km of a canal.[2]

[1] Schultz (1953) referred to this phenomenon as the "food problem". The same mechanism lies at the heart of more recent work, which relies on non-homothetic preferences as the main driver of structural transformation (Gollin et al., 2002, 2007; Alvarez-Cuadrado and Poschke, 2011; Comin et al., 2021). The link between agricultural productivity growth and structural change also emerges in other models where productivity growth leads to endogenous changes in the relative price of agricultural goods (Ngai and Pissarides, 2007).

[2] We study India's network of major and medium canals, for which data is maintained by the national Ministry of Water Resources. Smaller surface irrigation projects, such as channels diverting water from village tanks (small

These canals are an ideal context for studying the long-run impacts of technical change in agriculture because (i) they cause change in agricultural productivity with sharp boundaries that are sustained for decades; and (ii) the majority of canals were built before 1980, allowing us to study the spatial equilibrium that has emerged in the long run. Other agricultural interventions tend to gradually diffuse across space, making it more difficult to study their effects in the long run.

Specifically, we ask how the agricultural productivity gains from canals affected development and structural transformation at the local level. The distinctive features of our analysis are that we study the long-run equilibrium, and we examine structural change at different geographic scales. Much of the prior work has focused on a single geographic level of analysis, such as the village or the county. But factor mobility depends on geographic scale and time horizon; for example, labor may be highly mobile across villages but less mobile across states or language regions. We study how canal irrigation has shaped the economy in the long run at the level of irrigated villages, connected rural areas, and in the regional urban economy.

This approach requires detailed, high-resolution data. We combine microdata from business and household censuses, administrative records, geospatial datasets, and satellite imagery to measure irrigation, agricultural activity, living standards, and non-farm economic activity for all of India's 600,000+ settlements (villages and towns). Our main outcomes were recorded in 2011–2013, over 40 years after the beginning of construction for the median canal and 30 years after the median canal was declared complete. Enough time has passed since canal construction that we are arguably observing the spatial equilibrium that has emerged in the long run.

We can think of canals as having effects at four different geographies: (i) direct effects in the settlements that they serve with surface irrigation; (ii) indirect effects in nearby unserved settlements; (iii) effects in regional urban markets; and (iv) diffuse effects across much broader geographies, such as the entire country or world. We use distinct identification strategies to measure effects (i), (ii), and (iii); like much of the literature on the effects of place-based policies, we are unable to provide empirical evidence on universal effects.

---

artificial reservoirs) or streams to farmers' fields are not included in this analysis.

To measure direct effects on canal-irrigated areas, we use a regression discontinuity design (RDD) that exploits the gravity-driven nature of canal water distribution, with elevation relative to the nearest canal as the running variable. At the local level, canal placement is determined by engineering constraints and topography, and water from canals only flows downhill, treating settlements topographically below the canal. Settlements a short distance away but only a few meters higher than the canal will thus experience little to no irrigation benefit and can serve as a control group for the irrigation treatment.

The RDD analysis tests for long-run differences between places that have direct access to canal irrigation and those that do not, but it does not account for spillovers into above-canal settlements. For example, in the presence of local labor or goods market linkages, untreated settlements close to canals could increase demand for both agricultural and non-agricultural labor. Alternately, canals could recharge regional groundwater tables, increasing access to pump irrigation in the control settlements above the canal.

To measure spillover effects, we compare settlements directly above canals to settlements that are in the same district but are more distant from canals. This gives us three types of settlements — (i) below-canal (or directly treated) settlements, which are directly treated with canal irrigation; (ii) above-canal (or indirectly treated) settlements, which are exposed only to the spillover effects of canals; and (iii) distant settlements (or untreated settlements). Naturally, this setup only allows us to measure spillovers that shrink with distance; if spillovers extend frictionlessly to the entire country or world, then they are impossible to measure, given the kind of cross-sectional data that is available. We use entropy balancing (Hainmueller, 2012) to re-weight the sample of distant and above-canal settlements, ensuring a comparison of settlements with similar distributions of natural characteristics (climate, topography, and agricultural potential). If there are important spillovers, the above-canal settlements will be substantially different from distant settlements, even though neither group is directly treated with canal irrigation.

Finally, we consider the possibility that industry growth arising from canals occurs through concentrated production clusters, which could be overlooked by the analysis above. To address this possibility, we draw on a hundred-year panel of town-level urban population — a variable that is consistently measured across time and is available at high resolution.[3] We use a difference-in-differences design that studies town growth before and after regional canals are built, following Callaway and Sant'Anna (2020).

The RDD analysis reveals sharply improved agricultural outcomes in the settlements directly treated by canals. Treatment settlements have more irrigated acreage, increased land under cultivation, a higher likelihood of growing water-intensive crops, and greater estimated yields.[4] The yield effects are observed almost entirely in the relatively dry winter (*rabi*) season: canals thus improve water access in a second cropping season but generate no significant differences during the summer monsoon (*kharif*) growing season, when rainfall is much more plentiful. There are no spillover effects on a range of agricultural outcomes: irrigation levels, yields, and land use in above-canal settlements are highly similar to those in more distant settlements. The sharp differences in agricultural outcomes between above-canal and below-canal settlements have been sustained over many decades, making them a useful natural experiment for studying what happens to the rest of the economy in the long run following a major increase in agricultural productivity.[5]

The agricultural changes brought about by canals cause substantial population growth in irrigated regions, but ultimately little local structural change. Below-canal settlements have sharply higher population density, with minimal spillovers into above-canal areas. Below-canal, above-canal, and distant settlements have highly similar shares of workers employed in manufacturing, services, and even in agro-processing. There is evidently an increased demand for labor, as evidenced by higher population density in canal-irrigated areas, but these highly agricultural settlements do not develop substantial non-farm sectors. The town panel analysis also finds concentrated population gains in proximate urban areas in the decades following canal construction. The net population movements are substantial in magnitude; a back-of-the-envelope calculation suggests that India's canal network has increased the population of canal villages by nearly 32 million people and added 5–9 million people to canal-region towns.

Canals have heterogeneous effects on living standards. Using small area estimates from household

[3] Population is, in fact, the only such variable that is available in the era of canal construction.

[4] In the absence of high resolution directly-measured yield data, we use a satellite-derived proxy that estimates biomass added in a village over the course of a growing season.

[5] Results are robust to a wide range of alternate specifications, including a regression discontinuity using distance to the officially designated command area boundary of the canal.

data (Elbers et al., 2003), we find that canals produce no significant consumption gains for the roughly 70% of rural households who own little to no land. In contrast, households in the higher quartiles of landholding show substantial increases in consumption in the directly-irrigated zone, with effects increasing in the size of land holdings. There are no consumption spillovers into above-canal settlements, suggesting the wealth effect is driven by higher returns to land.

We interpret our results in the context of a multi-sector, multi-location model that is closely related to Matsuyama (1992) and Bustos et al. (2016) but captures two key features of our context. First, we model labor as immobile across space in the short run and fully mobile in the long run. Second, we assume that towns have productivity advantages in non-farm work relative to villages. Our model highlights several insights from the empirical results. In the long-run spatial equilibrium, increased demand for labor is met by an increase in the number of laborers, eliminating differences in wages across space. Workers still benefit, but the gains are spread across a large linked labor market, such that the local effect of any one canal is very small. Returns to land, the fixed factor, remain higher – even in the long run. Structural transformation occurs entirely through the growth of towns, rather than through the relative growth of the non-farm sector in rural areas.

This paper extends a substantial literature linking technical change in agriculture to structural change. Our results can help to tie together Foster and Rosenzweig (1996, 2004b), who found in a panel of Indian villages that the Green Revolution raised wages but inhibited local industrialization, and Bustos et al. (2016), who found in Brazilian municipalities that innovations in soybean and maize had divergent effects on structural change that depended on the changing factor intensity of agriculture. These papers operate on different geographic scales. Brazilian municipalities are ten times larger than Indian villages and incorporate regional towns; our paper measures outcomes separately at both of these scales. Our results show that, in the long run, net population movement is a central response to agricultural productivity gains and a key mechanism for structural change. Both villages and towns gained population, but with little change in their non-farm production structure.[6]

[6]Our results also recall Bustos et al. (2020), who found that agricultural productivity gains drove urbanization through the flow of land rents to cities. An example of this capital channel was discussed at length in the context of colonial Bengal in Bose et al. (1993).

Canal-irrigated villages grew due to increased labor demand in agriculture, while the downstream effects of those agricultural gains were concentrated in towns. Our results parallel the findings of Hornbeck and Keskin (2015), who found in the U.S. that the tapping of the Ogallala Aquifer increased agricultural productivity but had little effect on local rural non-agricultural activity in the long run.

Our results also echo Foster and Rosenzweig (2004a), who argued that agricultural productivity shocks have substantially different effects on landowners and the landless, consistent with our findings. There is also a large body of evidence on responses to transient agricultural productivity shocks due to weather (Adhvaryu et al., 2013; Colmer, 2021). Emerick (2018) and Santangelo (2019) in particular find that non-tradable employment increases in districts experiencing positive agricultural productivity shocks, consistent with our model of demand-driven structural change. This paper speaks less to this literature on transient shocks because we study how people adjust to large, permanent changes in agricultural productivity.

Finally, we contribute to the literature on how labor flows respond to economic shocks. A rich body of research documents how economic shocks can result in both temporary and permanent migration in both high- and low-income countries (Greenstone et al., 2010; Allcott and Keniston, 2018; Imbert and Papp, 2020). Much of the recent empirical work has aimed to study competition for workers between the farm and non-farm sectors in models that shut down the labor mobility channel. This is partly for the reason that mobility is typically lower in the short- to medium-run periods examined in prior studies. Indeed, in an extension of their main results, Bustos et al. (2016) find that about one-third of the shift out of agricultural employment in soybean areas occurred via migration, over only a 10-year sample period. Our longer-run analysis suggests that the movement of people can be the primary adjustment channel to agricultural change. Indeed, the very nature of structural transformation around the world has involved the movement of billions of people from farms to cities, sometimes across large distances.[7]

Our results further highlight the high barriers to rural industrialization. Asher and Novosad

[7]While there is a widespread idea in the literature that permanent migration in India is rare, this idea is focused on the set of rural men who migrate for work. Over 25% of women have changed residence at least once in their lives, and lifetime migration rates for men approach 15% (Kone et al., 2018). Since we only observe population density in the present, we cannot distinguish migration from other mechanisms of population change, but we find suggestive evidence against fertility and mortality as primary channels for the population density effects of canals.

(2020) and Burlig and Preonas (2021) find that major investments in rural roads and electrification respectively have generated limited effects on non-farm activity in India.[8] Faber (2014) finds that highway construction through peripheral areas in China in fact caused deindustrialization. These papers suggest that while infrastructure investments in rural areas may improve well-being, they often do *not* cause substantial changes in *in situ* non-farm opportunities. Our results are also consistent with long-run evidence that the Green Revolution had substantial positive effects on structural change (Gollin et al., 2021); our analysis suggests that this process may have been driven by the growth of urban areas, rather than in the rural areas directly experiencing agricultural productivity gains.

Finally, our work adds to a growing literature estimating the impacts of access to irrigation. Sekhri (2014) shows that access to groundwater irrigation lowers poverty. Blakeslee et al. (2021b) find that the drying up of wells induces agricultural households to provide more non-farm labor. Jones et al. (2020) study canal irrigation in Rwanda using an elevation-based RDD, finding that labor market frictions limit the returns and thus lower adoption of irrigation. In a concurrent paper, similar to this one, Blakeslee et al. (2021a) study canals in India using an RDD exploiting the official canal command area boundary. While they find similar reduced form effects on population density and (lack of) structural change in canal-irrigated villages, their analysis does not consider spillovers and is focused primarily on evaluating canals as infrastructure investments, rather than as drivers of long-run agricultural change.[9] We extend this literature by providing contrasting evidence on long-run impacts for directly treated areas and indirect effects on proximate rural and urban areas.

The rest of the paper proceeds as follows: Section 2 provides context on the role of canals in Indian agriculture. Section 3 develops a model of how canals may affect economic activities at different geographic levels and time horizons. Sections 4 and 5 describe the data and our multiple empirical strategies. Section 6 presents results, Section 7 discusses interpretation, and . Section 8 concludes.

[8] Asher and Novosad (2020) find that the main impact of roads is to provide access to non-agricultural labor markets outside the village. This result is suggested by our model, where towns have productivity advantages for non-farm work.

[9] We show our elevation-based RDD results are consistent with those from this approach.

2 Context

As a semi-arid region with a highly variable monsoon climate, South Asia has long depended on irrigation for its agricultural productivity. For much of history, this has primarily involved gravity flow surface irrigation through canals of various types. At the end of the 19th century, India's had 12 million hectares of irrigated land — four times more than the United States and six times more than Egypt (Shah, 2011). The British oversaw the construction of vast canal networks, often privately funded and yielding high returns, until the end of the Raj in 1947. Canals were used to divert water from India's major rivers to its arid regions, where they facilitated settlement of otherwise uninhabitable land, such as with the Punjab Canal Colonies (Douie, 1914). After gaining independence, the Government of India prioritized canal-building as it sought to avoid mass hunger during a period of high population growth (Mukherji, 2016). Later, canals were built to provide irrigation for the input-intensive high-yielding varieties of food crops that powered India's Green Revolution.

While groundwater eclipsed canals as India's preeminent source of irrigation by the 1970s (Shah, 2011), surface irrigation remains critical to the livelihood of millions of farmers across India. In recognition of the importance of canals, the central government launched the Accelerated Irrigation Benefit Program (AIBP) in 1997, which spent more than $7.5 billion on improvement and completion of large-scale irrigation projects (Shah, 2011). According to the most recent estimates, canals still account for approximately one-fourth of the net irrigated area in India (Jain et al., 2019), although estimates vary according to the methodology.

Figure 1 shows the distribution of official completion dates of India's major and medium canals according to India's Ministry of Water Resources. A caveat to this figure is that the official "completion date" is updated if a canal undergoes a substantial renovation. As a result, the older dates (when India had fewer canals) mostly represent new canals, whereas many of the recent dates in fact reflect rehabilitation projects on canals built several decades earlier.

Major canals are defined as serving 10,000 or more hectares while medium canals serve areas 2,000–10,000 hectares. Canals serving 2,000 hectares or less are termed minor canals and are not included in this study. Construction rates increased following India's independence in 1947, although post-independence canals are generally shorter than those constructed under the British Raj in the 19th and early 20th century. By 2011, 51% of India's 600,000 settlements were within 10 km of a major or medium irrigation canal, with a median canal construction start year of 1972 and completion year of 1980. Given that our primary outcomes are measured in 2011–2013, the canals in our study are typically at least thirty years old.

3 Model

Our theoretical framework builds on a substantial literature modeling the effects of agricultural productivity change on the non-farm sector (Johnston and Mellor, 1961; Matsuyama, 1992; Foster and Rosenzweig, 1996, 2007; Bustos et al., 2016). Early models in this literature tended to predict that an increase in agricultural productivity (crucially, in a closed economy) would lead to a decline in the relative price of agricultural goods. This in turn lowers the returns to inputs used in this sector and induces a movement of productive resources into non-agricultural sectors. This mechanism lies at the heart of Johnston and Mellor (1961), Ranis and Fei (1961), and Jorgenson (1961), as well as subsequent papers (Eswaran and Kotwal, 1993; Gollin et al., 2002; Restuccia et al., 2008; Alvarez-Cuadrado and Poschke, 2011).

However, the relationship between agricultural productivity gains and structural transformation has been shown to depend on assumptions relating to the openness to trade (Matsuyama, 1992), the substitutability of agricultural and non-agricultural goods (Ngai and Pissarides, 2007), the factor intensity of technological change (Bustos et al., 2016), and capital mobility (Foster and Rosenzweig, 2007; Bustos et al., 2020), among others. We deviate from existing models in the literature in two key dimensions, which reflect our empirical context. First, we model an economy in which labor flows freely across space in the long run, but not in the short run. This offers a contrast to many models that allow for labor mobility across sectors but not across locations.[10] Second, we allow for spatial variation in non-agricultural productivity, such that larger settlements have a productivity advantage in the production of non-agricultural goods.

[10] In this respect, we are most closely related to Foster and Rosenzweig (2007), who recognize the importance of factor mobility – although in their case, the mobile factor is capital rather than labor.

We model India's rural economy as a large number of predominantly local sub-economies that are embedded in a larger national economy; we focus on one such representative region. Each rural region features an expanse of agricultural land, divided into villages, typically with a larger market town that serves as an economic center.[11] Agricultural land is in general privately owned and managed. Most farms are small (Foster and Rosenzweig, 2017), and many landowners work on their own land. Farms may also hire labor from a large pool of landless workers. These observed features of the data give shape to our conceptual framework.

3.1 Model Setup

The model focuses on a rural region that is comprised of a single town and a set of surrounding villages. Let V denote the number of villages, and let v_i denote the i^{th} village, $i=1,2,...,V$. In what follows, we simplify to an environment where $V=2$. We designate the town as settlement $i=0$, and the two villages as $i\in\{1,2\}$. The region is embedded in a national economy, which is comparatively large.

The economy produces two goods: an agricultural good a that is traded beyond the region and a non-agricultural good c that is costlessly traded within the region and non-tradable beyond the region. The non-tradable good might correspond to services, such as haircuts; but it could also represent manufactured goods with low value per unit transport costs, such as bricks.[12]

Individuals consume the two locally produced goods, a and c, as well as a third good m: a traded non-agricultural good that is only available from the rest of the economy. This represents a class of goods that requires production capabilities that are not available within the rural economy (e.g., mobile phones) or perhaps some raw materials that are also unavailable locally (e.g., refined petroleum products). The rural region pays for these "imported" goods through "exports" of its agricultural production. We limit our analysis to the case where this economy is a net exporter of agricultural goods.

[11] The villages that surround each market town are mostly small; in 2011, the median village population in India was 844. Most villagers work in agriculture. (The median number proportion of non-farm jobs in a village is 5 per 100 adults according to the 2013 Economic Census and 2011 Population Census).

[12] This reflects the fact that in much of non-urban India, villages and towns produce a mix of non-tradable services (e.g., wholesale and retail trade, food service entertainment, government administration and public sector work, construction, repair services, and personal care) and relatively non-tradable manufacturing (e.g., brick making, metal fabrication, and carpentry). The vast majority of manufacturing firms in India have under five employees, and are thus unlikely to be serving a very large market.

We consider three periods. In the initial period, the region is in a long-run spatial equilibrium with the rest of the country. Following the initial period, a canal is built that raises agricultural productivity in village 1 but not village 2. During the second period, which describes the short run, labor is mobile across sectors and settlements within the region, but not between the region and the rest of the country. In the third period, which we call the long run, labor is also mobile across regions.

3.1.1 Preferences and utility

The representative consumer has preferences over the three consumption goods. These preferences can be represented by a log linear utility function:

$$u(a,c,m) = \alpha \log a + \beta \log c + (1-\alpha-\beta)\log m \tag{3.1}$$

For simplicity, we use homothetic preferences; this is convenient for aggregation and does not require us to address issues related to (for example) the distribution of land across households.

3.1.2 Production and trade

The agricultural good is produced on the village land, and the non-agricultural good can be produced either in villages or in towns.

Each of the region's villages has an endowment of land (L_i) and labor (N_i), while the town has only labor (N_0). The regional economy has a labor force of N people, where N_i is the labor force of village v_i and N_0 is the labor force of the town. Thus, $\sum_i N_i = N$. The supply of land is fixed in all periods, while the total regional labor force N is fixed only in the short-run following canal construction. For simplicity, we assume that all land in the region is held by a single landowner, who resides in the town and receives all land rents. All individuals supply one unit of labor to the market, inelastically.[13]

The agricultural technology is Cobb-Douglas, $Y_{a_i} = A_i N_{a_i}^{\theta} L_{a_i}^{1-\theta}$, where A_i represents agricultural productivity in village i, N_{a_i} and L_{a_i} denote land and labor in agriculture in settlement i, $0 < \theta < 1$, and $i \in \{1,2\}$. The non-agricultural good is produced with a technology that is linear in labor:

[13]Because every individual in this model is a worker, including the landowner, we use the terms labor force and population interchangeably.

$Y_{c_i} = C_i N_{c_i}$, $i \in \{0,1,2\}$, where C_i is the non-agricultural productivity term. We assume that due to natural advantage or agglomeration economies, the town has the highest C_i in the region. Recall that the traded good m is consumed but not produced within the region.

Since both the agricultural good and the manufactured good are traded frictionlessly with the rest of the economy, the representative region is a price taker for these two goods. The relative price p_m is the price of this imported manufactured good in terms of agricultural goods, which are the numeraire. The price of the non-tradable good p_c is determined endogenously in the region and depends on the productivity level for non-tradables. Because labor always moves frictionlessly across settlements and sectors within the region, there is a single regional wage w.

3.2 Equilibrium

An equilibrium consists of an allocation of labor across settlements and sectors (N_0, N_{a1}, N_{c1}, N_{a2}, N_{c2}), prices (p_c, p_m), and the wage w.

Because the non-tradable good is frictionlessly traded within the region, and because the production technology is linear in labor, the non-tradable good is produced in all periods only in the settlement with the highest productivity level; by construction, this is always the town. Thus, the non-tradable good will be produced only in the town and because the town has no land, it will produce only the non-tradable good. We can dispense with location subscripts and define total regional output of the non-tradable good as $Y_c = Y_0 = C_0 N_0$. Due to the zero profit condition, the non-tradable price is fixed at $p_c = \frac{w}{C_0}$.

Because the economy faces no externalities or market imperfections, and because production is fully competitive, the first and second welfare theorems hold, and we can solve the social planner's problem to arrive at the same equilibrium allocations that would obtain in a competitive equilibrium. Moreover, since preferences are homothetic, we can focus on the problem of a representative consumer who receives the average consumption allocation.

3.2.1 Long-run equilibrium

In the long-run spatial equilibrium where labor is fully mobile across regions, workers have the same utility ($\bar{u}$) everywhere. Because the region is a price taker for both goods a and m, utility is fully determined by the wage and the price of the non-tradable good c. With all variables that affect the local wage thus fixed, we can see that the long-run wage w_{LR} does not depend on the agricultural productivity of either village.

From the consumer's problem, we know that the budget share for the non-tradable good is given by the corresponding elasticity β in the Cobb-Douglas utility function. Total income for the regional economy is the value of output. Since villages produce only a and the town produces only c, and taking the agricultural good as the numeraire, this gives $Y = Y_a + p_c Y_c = Y_a + w_{LR} N_c$, where total agricultural output $Y_a = Y_{a_1} + Y_{a_2}$. Expenditure on the non-tradable good is thus $\beta(Y_a + w_{LR} N_c)$, and production value is $p_c Y_c = w_{LR} N_c$. This gives the following condition for non-tradable employment:

$$N_0 = N_c = \left(\frac{1}{w_{LR}}\right)\left(\frac{\beta}{1-\beta}\right) Y_a. \tag{3.2}$$

Agricultural production: Agricultural employment and output in each village are pinned down by the price of the agricultural good and the regional wage w_{LR}. Given the Cobb-Douglas production technology, agricultural employment in each village is given by:

$$N_{ai} = \left(\frac{\theta A_i}{w_{LR}}\right)^{\frac{1}{1-\theta}} L_i,\ i = 1,2. \tag{3.3}$$

This in turn gives output for each settlement of:

$$Y_{ai} = A_i^{\frac{1}{1-\theta}} \left(\frac{\theta}{w_{LR}}\right)^{\frac{\theta}{1-\theta}} L_i,\ i = 1,2. \tag{3.4}$$

The total agricultural output is thus:

$$Y_a = Y_{a1} + Y_{a2} = \left(\frac{\theta}{w_{LR}}\right)^{\frac{\theta}{1-\theta}} \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2\right]. \tag{3.5}$$

Combining equations 3.2 with 3.5, we can solve for the non-tradable labor force:

$$N_0 = w_{LR}^{\frac{1}{\theta-1}} \theta^{\frac{\theta}{1-\theta}} \left(\frac{\beta}{1-\beta}\right) \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2\right]. \tag{3.6}$$

The non-tradable output is trivially $Y_c = C_0 N_0$, or:

$$Y_c = Y_0 = C_0 w_{LR}^{\frac{1}{\theta-1}} \theta^{\frac{\theta}{1-\theta}} \left(\frac{\beta}{1-\beta}\right) \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2\right]. \tag{3.7}$$

This set of equations fully specifies the long-run equilibrium. Total population is given by $N = N_0 + N_1 + N_2$.

Comparative statics: Canal construction raises agricultural productivity in village 1 (A_1). This increases the demand for labor, causing the population of village 1 to increase until the marginal product of labor is brought back to to the long-run wage w_{LR}. There is no effect on the population of village 2, as each village's equilibrium population depends only on the wage, its own agricultural productivity, and its endowment of land; none of these are affected by a change in agricultural productivity in village 1. Land rents increase in village 1 only. The increase in population, along with higher land rents, raises demand for the non-tradable good and thus the population of the town. This implies a higher overall population in the region – an inflow of workers from outside the region due to the construction of the canal.

3.2.2 Short-run equilibrium:

We now examine the short-run equilibrium following canal construction. There is no population flow between the region and the rest of the country, but labor markets clear within the region. The region's population is fixed at the level of the initial period. Because labor does not flow across regions, the wage is no longer pinned down by the national reservation utility $\bar{u}$ and can deviate from w_{LR}.

Let N^0 be the initial long-run equilibrium population prior to the construction of the canal in village 1. The equilibrium wage, conditional on this population level, is determined by the labor market clearing condition: $N_0 + N_1 + N_2 = N^0$. Plugging in the values of N_0, N_1, and N_2 derived above, we get the following expression:

$$w^{\frac{1}{\theta-1}}\theta^{\frac{\theta}{1-\theta}}\left(\frac{\beta}{1-\beta}\right)\left[A_1^{\frac{1}{1-\theta}}L_1+A_2^{\frac{1}{1-\theta}}L_2\right]+\left(\frac{\theta A_1}{w}\right)^{\frac{1}{1-\theta}}L_1+\left(\frac{\theta A_2}{w}\right)^{\frac{1}{1-\theta}}L_2=N^0.$$

This simplifies to:

$$w={N^0}^{\theta-1}\theta\left[\left(\frac{1}{\theta}\right)\left(\frac{\beta}{1-\beta}\right)+1\right]^{1-\theta}\left[A_1^{\frac{1}{1-\theta}}L_1+A_2^{\frac{1}{1-\theta}}L_2\right]^{1-\theta} \tag{3.8}$$

The change in the wage is given by

$$\frac{\partial w}{\partial A_1}=\frac{\left[\left(\frac{1}{\theta}\right)\left(\frac{\beta}{1-\beta}\right)+1\right]^{1-\theta}\theta {N^0}^{\theta-1}A_1^{\frac{\theta}{1-\theta}}L_1}{\left[\sum_{i=1}^{2}\left(A_i^{\frac{1}{1-\theta}}L_i\right)\right]^{\theta}}.$$

This partial derivative is unambiguously positive; the increase in agricultural productivity in village 1 drives up the regional wage. The impact on population in village 2 is unambiguously negative; the higher wage reduces agricultural employment and thus output in this village. The effect on the non-tradable sector and thus the town population depends on parameter values. Intuitively, the higher wage has a direct crowd-out effect on employment in the non-tradable sector, but an indirect crowd-in effect through increased regional demand for the non-tradable good. The effect on village 1 is likewise ambiguous, as the increased wage and increased agricultural productivity have countervailing effects.

The model illustrates the two major contributions of this paper. First, the long-run impacts of agricultural productivity gains vary with geography: treated villages gain agricultural workers while non-agricultural growth occurs in higher productivity urban areas. Second, the long-run impacts of agricultural productivity shocks may be different from the short-run impacts due to the mobility of labor. Because the irrigation canals that we study were built so long before the high-resolution data needed to study their effects, the empirical analysis describes the long-run equilibrium after canal construction.

4 Data

To study how canal irrigation affects local economic outcomes, we assemble recent high-resolution data on the universe of firms, households, and settlements in India, building on data from the SHRUG (Asher et al., 2021). Because the reclassification of rural villages into urban towns is an endogenous outcome driven by population density and administrative discretion, we combine villages and towns into a single dataset; we use the term 'settlements' to include both categories. The dataset covers 590,000 settlements (8,000 are towns; the rest are villages), which are nested in 5,000 subdistricts and 700 districts.

The 2011 Population Census provides demographic variables and data on cultivated and irrigated land area in every village in India. The Census also records the three main crops grown in each village, from which we create an indicator for villages that grow a water-intensive crop (cotton, sugarcane, or rice).[14] Since settlements are heterogeneous in size, our preferred measure of population is density, which we define as inhabitants per km^2.[15]

The 2012 Socioeconomic and Caste Census (SECC) is an asset census that was undertaken in all of India to determine eligibility for means-tested programs (Asher and Novosad, 2020). From SECC microdata, we generate the share of adults aged 20–65 who have completed primary, middle, and secondary school, as well as predicted consumption per capita using small area estimation based on the income and asset variables in the SECC.[16] Because the SECC is recorded at the household level, we can calculate these outcomes separately for landowners and landless households.

The 2013 Economic Census is a complete enumeration of all non-farm economic establishments in India, which we use to measure non-agricultural economic activity for each settlement. We calculate employment as a share of the 2011 Census adult population.[17] We use the National Industrial

[14] Population Census data on agricultural outcomes are available only in villages; analysis of these outcomes therefore excludes towns.

[15] We calculate population density as settlement population divided by the area of the settlement GIS polygon shape (in km^2) as opposed to the noisier area reported in the Population Census. See below for description of GIS data.

[16] The latter follows the methodology of Elbers et al. (2003) and is described in detail in Asher et al. (2021). For a secondary measure of educational attainment, we use the settlement literacy rate from the Population Census.

[17] As the Population Census only reports age-disaggregated numbers for the population aged 0–6, we estimate the population aged 0–17 by multiplying the 0–6 population by 18/7. We then subtract the estimated 0–17 age group from the total population to get the adult population. This calculation reflects the fact that the Indian population pyramid in 2013 is close to uniform for ages 0–30.

Classification codes of firms in the Economic Census to calculate the share of the adult population employed in manufacturing, services, and agro-processing.[18]

In the absence of directly-measured settlement-level agricultural productivity data, we use the Enhanced Vegetation Index (EVI), a satellite-derived measure of biomass that has been widely used as a proxy for agricultural productivity (Wardlow and Egbert, 2010; Kouadio et al., 2014; Son et al., 2014). We calculate productivity for both the monsoon (*kharif*) season, late May through early October, and winter (*rabi*) season, late December through late March (Selvaraju, 2003). For each season, we define productivity by subtracting the mean of the first six weeks of the season from the maximum EVI value reached during the entire season following Rasmussen (1997) and Labus et al. (2002). This measure has better prediction accuracy for yield than a raw biomass measure, as the latter may pick up forest land, which registers as high biomass, but does not change as much as agricultural land during the cropping season. We calculate the mean of this measure for years 2011–13 (corresponding to our other outcome datasets), and log transform it to address outliers and simplify interpretation.[19]

Spatial data on canals and their command areas comes from the Ministry of Water Resources. The India Water Resources Information System (WRIS), a part of the Management Information System of Water Resources Projects of the Central Water Commission in India, provides geospatial data on canals and their command areas.[20] The command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal, extending out from the canal and ending at a boundary that is determined by a combination of canal flow, terrain, and soil type. The WRIS provides dates of canal construction and completion; however, our research on individual canals suggests that recent start and end dates in WRIS often represent canal rehabilitation efforts, rather than new canal construction.[21] It is therefore challenging to identify exact construction dates of recent

[18] Manufacturing employment contains NIC 2-digit codes 10–35 (excluding only the 3-digit code 131) while services contains NIC 2-digit codes 36–93 and 131. Agro-processing is defined as a subset of manufacturing employment codes, specifically NIC codes 10 and 12.

[19] We find similar results if we use different years (which is expected, given that we are studying equilibrium effects of canals) or EVI levels rather than logs. See Asher and Novosad (2020) for more details on construction of the EVI measure.

[20] The database can be found at `https://indiawris.gov.in/wris/`.

[21] The WRIS database often reports construction dates only in terms of a 5-year planning period, meaning dates are only known within a 5-year window. We augmented and verified dates from the database by manually searching for canal construction dates reported in government documents, news articles, ministry reports, and academic papers.

canals. Older construction dates appear to be more credible, as canal investments in the independence period and earlier were more often new canals rather than maintenance of existing infrastructure.

Using settlement polygon GIS data from ML Infomap, we extract the distribution of elevation in each settlement from Shuttle Radar Topography Mission (SRTM) raster data. Following Riley et al. (1999) and Nunn and Puga (2012), we calculate the ruggedness of a location's topography using the Terrain Ruggedness Index (TRI); TRI measures ruggedness as the average square difference in elevation between a pixel and its eight surrounding pixels. We take the average TRI value across all pixels in a settlement to characterize ruggedness. Using these same data, we compute the distance from every settlement centroid to the nearest canal, command area, river, and coast.

Using the same settlement polygons, we extract 10-day rainfall values from the Climate Hazards Center InfraRed Precipitation with Station (CHIRPS) dataset (Funk et al., 2014). We calculate the average of total annual rainfall from 2010-2014 as our rainfall measure. Similarly, we extract monthly maximum daily temperature for each settlement from the Climate Hazards Center Infrared Temperature with Stations (CHIRTS) dataset (Funk et al., 2019), then compute the average maximum monthly temperature over the 2010-2014 time period as our temperature measure.

5 Empirical Strategy

Testing for the long-run impacts of increasing agricultural productivity is challenging for two reasons. First, the placement of canals is endogenous: large, costly infrastructure investments tend to be targeted to areas that are politically favored and have high returns to irrigation. Second, canals can have different effects at different geographic scales. To overcome these challenges, we use three empirical strategies, each of which isolates a different aspect of the effect of canals. To estimate the direct effects on locations receiving increases in agricultural productivity, we exploit the gravitational nature of canal irrigation, which creates arbitrary differences in irrigation availability in proximate settlements directly above and below the canal. To test for the presence of spillovers into nearby untreated locations, we use a matching estimator to compare both above- and below-canal settlements to settlements that have similar geophysical characteristics but are further away from canals. Finally, to test for effects on regional urban growth, we use a hundred-year panel of town populations and a difference-in-differences estimator.

5.1 Regression Discontinuity Estimates of the Direct Effects of Canals

Canals provide water to fields through a system of gravity-driven secondary canals, trenches, and pipes. Because water delivery depends physically on gravity, fields must be at a lower elevation than a canal in order to be irrigated with canal water; settlements above the canal will not benefit directly. Our main identification strategy compares settlements close to canals with elevations that put them either just above or just below the threshold that would give them access to canal water. For this analysis, below-canal settlements are treated by canals and above-canal settlements serve as controls.

A settlement polygon is characterized in the data by a set of pixels with a distribution of elevation values. We define the polygon elevation as the 5$^{\text{th}}$ percentile of the polygon pixel distribution; this value strongly predicts the difference in canal irrigation between treatment and control areas (see Appendix Figure A1).[22] For each settlement, we also calculate the elevation of the canal at its nearest point to the settlement.

Equation 5.1 describes the regression discontinuity design (RDD) specification, following Imbens and Lemieux (2008) and Gelman and Imbens (2019):

$$\begin{gathered} y_{i,s} = \beta_0 + \beta_1 1\{REL\_ELEV_{i,s} < 0\} + \beta_2 REL\_ELEV_{i,s} + \\ \beta_3 REL\_ELEV_{i,s} * 1\{REL\_ELEV_{i,s} > 0\} + \beta_4 X_{i,s} + \nu_s + \epsilon_{i,s}, \end{gathered} \tag{5.1}$$

where $y_{i,s}$ is an outcome in settlement i and subdistrict s and $REL\_ELEV_{i,s}$ is settlement elevation minus canal elevation (such that a negative value means that the settlement lies below the canal, and thus can receive its water), and $X_{i,s}$ is a vector of geophysical controls (ruggedness, mean annual rainfall, maximum annual temperature, distance to the nearest river, distance to the coast, and the GAEZ crop suitability measure for irrigated rice and wheat).[23] ν_s is a subdistrict fixed effect, which

[22] Results are similar if we use the 25$^{\text{th}}$ percentile or median elevation to define above/below canal thresholds (Appendix Tables A3, A4, A5, and A6). We chose the 5$^{\text{th}}$ percentile in order to have a control group with close to zero canal irrigation; when we estimate spillover effects below, interpretation is most straightforward if the above-canal group experiences no direct treatment by canal water.

[23] We use these as proxies of agricultural fertility and potential returns to irrigation, which could have hypothetically guided canal placement. As agriculture in India tends to use some inputs but not nearly as much as rich countries, we use the intermediate input variables from the FAO GAEZ. We do not include any socioeconomic controls, because they are available at the settlement level only after 1990, by which time they are plausibly affected by canals.

restricts our above/below canal comparison to settlements in the same subdistrict. A subdistrict consists of approximately 100 settlements, with a total population of about 250,000 people. Standard errors are clustered at the subdistrict level to account for spatial correlation. In the absence of spillovers to untreated settlements, the effect of canal irrigation is captured by β_1, which is the difference in outcomes between settlements just below and just above the canal. Appendix Figure A2 shows a map of a single district, along with its canal network, elevation profile, and the first stage RDD graph showing the share of land irrigated by canal.

The main analysis sample includes settlements less than 10km of distance and 50m of vertical elevation from the nearest canal.[24] As our outcome data is from 2011 onwards, we exclude from our analysis sample any settlements whose closest canal is listed as incomplete as of 2011. We limit the sample to subdistricts that have at least one settlement in the treatment group and one settlement in the control group. Settlements with elevation very close to the treatment threshold have an ambiguous treatment status — for example, a settlement could have some of its land above the canal (and thus not treatable with canal water) and some of its land below the canal (and thus treatable). Inclusion of these settlements would bias RDD estimates toward zero; we therefore exclude a "donut hole" of settlements within 2.5m in elevation of the nearest canal in either direction. Finally, to avoid comparing lowland irrigated areas with rugged hilly areas, we impose a balance restriction on the terrain ruggedness index. We allow a maximum 25% difference in mean ruggedness between below-canal and above-canal settlements in a given subdistrict; if the percent difference is greater, the entire subdistrict is dropped from the sample. Table 1 shows the sample size and mean values for all variables used in our analysis after each stage of the sample selection. We use the ruggedness-balanced sample (Column 4) for our primary analysis, but show robustness in the Appendix to alternate sample definitions. The ruggedness-balanced analysis sample is representative of the universe of settlements in India on most dimensions. Around half of agricultural land is irrigated; about 60% of village land is dedicated to agriculture; there is approximately 1 non-farm job for every 10 adults; and just under half of adults have completed primary school.

[24] It is rare that villages further than 10km from a major or medium canal branch show economically meaningful access to canal irrigation, even if they are below the elevation of the canal.

RDD validity requires that there are no pre-treatment differences at the threshold between above- and below-canal settlements. Since canal infrastructure in India was built throughout the 19th and 20th centuries, and treatment status is determined at the settlement level, there are no high-resolution socioeconomic or agricultural data available to test this assumption. However, we can test for differences in time-invariant geophysical measures, which could proxy for natural advantages that might have affected canal placement and economic outcomes. Table 2 shows estimates of Equation 5.1 on geophysical fundamentals (with the specific outcome excluded from $X_{i,s}$ in each regression), demonstrating that there are no significant differences between above- and below-canal settlements in ruggedness, distance to coast, average annual rainfall, or crop suitability for rice or wheat. We do find small imbalances on temperature and distance to rivers. The temperature difference is tiny in magnitude and would if anything make canal areas less productive. Canal areas are 10% further from rivers; this makes sense given that they aim to provide irrigation water where there is none. We control for all of these geophysical variables in all of the regressions below.

As a robustness check, we use a secondary regression discontinuity design that compares settlements just inside and just outside of the canal command area.[25] We define the running variable as the distance between settlement centroid and command area boundary, defining it negatively inside the command area.[26] The estimation is otherwise similar to that above, but we additionally divide each command area boundary into 10km segments and include a fixed effect for each segment, ensuring that we are comparing settlements across the same stretch of each command area. Standard errors are clustered by these segments. This strategy exploits the variation in the xy-plane, whereas the primary (relative elevation) strategy exploits variation in the z-axis. The identifying assumption is that settlements just inside and just outside the command area boundary would have similar outcomes if the canal had not been built. While the command area definition may exploit finer details of local topography, we prefer the relative elevation strategy, as boundaries of command areas may be subject to some discretion by officials, who might have incentives to mark one settlement or another

[25] This is similar in design to the strategy used in concurrent work by Blakeslee et al. (2021a). Recall that the command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal.

[26] The analysis sample contains settlements within 25km of the command area boundary, and the donut hole excludes those within 2.5km of the boundary. Results are similar with different exclusion criteria.

as within the command area.[27] We test for balance with this command area boundary strategy in Table A2, finding no evidence for any imbalance, apart from a very small difference in temperature.

5.2 Testing for spillovers into above-canal areas

The regression discontinuity design exploits arbitrary differences in access to canal water in proximate above- and below-canal settlements. Given that we are estimating long-run effects of canals, spillovers in such a small geographic area are a distinct possibility. For example, if above- and below-canal settlements are part of integrated labor markets (as they are in the model), then the labor market effects of canal irrigation would be expected to diffuse across the treatment boundary. If labor mobility was sufficiently high, we could estimate zero differences between these areas in the RDD analysis even in the presence of substantial labor market effects of canals. More directly, canals could recharge underground aquifers, improving access to pumped groundwater in above-canal areas.

To test for spillover effects, we define an alternative sample of control locations: distant settlements within each district, which lie at least 15km from the nearest canal but have similar geophysical characteristics. These settlements are at minimum 5 km further from the nearest canal than any treatment or control settlement in the RDD sample. By comparing settlements directly above the canal to these more distant settlements, we can test for the spillover effects of canals.

This strategy is predicated on the assumption that any mechanism driving spillovers is likely to decay with distance from treated areas. If spillovers do not decay over distance, they are more difficult to measure. For example, if landless labor were perfectly mobile across all of India, then a new canal could have a small positive impact on wages in the entire country, but there would be no control group against which such an effect could be measured. While we cannot rule out universal effects like these, our empirical design will identify the existence of spillovers as long as they have a non-zero gradient in distance. These spillovers point estimates will be biased downward to the extent that the effects of canals extend into the distant control group.

We use entropy balancing (Hainmueller, 2012) to assign weights to settlements so that the

[27] In practice, many of the treatment and control areas are defined similarly under the two strategies, since the command area is mechanically below the canal elevation.

distributions (first, second, and third moments) of all geophysical variables in distant, above-canal, and below-canal settlements are similar. Entropy balancing is a useful matching method because it does not impose functional form assumptions on propensity weights and thus achieves better balance than propensity-score matching.[28] Following the literature, we enforce common support by dropping outliers (the top and bottom 2.5% for each of the matching variables). As an alternate strategy, we use coarsened exact matching (Iacus et al., 2012) to similarly define and weight a matched sample of above-canal, below-canal, and distant settlements. This method discretizes the continuous geophysical variables into bins and matches settlements exactly on the coarsened variables. Distant settlements that do not perfectly match any above-canal settlements on these binned geophysical characteristics are discarded, and the remaining distant settlements are re-weighted to match the characteristics of the above-canal settlements. We test for spillovers using the following estimating equation:

$$y_{i,d}=\gamma_0+\gamma_1 1\{ABOVE\_CANAL_{i,d}\}+\gamma_2 1\{BELOW\_CANAL_{i,d}\}+X_{i,d}+\nu_d+\epsilon_{i,d}, \tag{5.2}$$

where above-canal and below-canal settlements are defined as in Section 5.1. Distant settlements are the omitted group. $X_{i,d}$ is the same vector of time-invariant geophysical controls as in the RDD specification above. To compare to more distant villages, we use a district fixed effect ν_d instead of the sub-district fixed effect in the RDD, and standard errors are clustered at the district level. The coefficient γ_1 describes the difference between above-canal settlements and distant settlements. If there are substantial spillovers from canal-irrigated areas into above-canal settlements, we expect γ_1 to be non-zero.

Note the difference between γ_2 here and the RDD estimate of β_1 from Equation 5.1. The RDD estimate describes the difference *at the threshold* between above- and below-canal settlements; γ_2 is the estimate of the average difference between below-canal settlements and distant settlements. If there are no spillovers, and there is no relationship between the RDD running variable (elevation) and the outcome, then we will find $\gamma_1=0$ and $\gamma_2=\beta_1$. In practice, the RDD estimator β_1 requires weaker assumptions for causal interpretation than γ_2 and is thus a better estimator of the direct effects of canal irrigation.

[28]See Athey and Imbens (2017) for more discussion matching methodologies, include entropy balancing. For recent examples of empirical work using entropy balancing, see Basri et al. (2021) and Guriev et al. (2021).

5.3 Town growth through time

Our model suggests that non-farm work may be concentrated in production clusters that have natural advantages or agglomeration economies. The empirical strategies thus far measure differences between canal-irrigated settlements, proximate non-irrigated settlements, and similar settlements farther away. Structural change that is concentrated in towns may not be captured by these tests for two reasons. First, whether a town is directly in the irrigation or spillover zone may be irrelevant for its prospects for non-farm work. Second, the spillovers analysis above estimates average effects and is not well suited to test for concentrated changes in a small number of towns in a sample mostly comprised of rural villages.

To test whether canals affect regional urbanization, we exploit variation in canal construction dates and examine whether town growth changes following the construction of regional canals. The available data (from the 2011 Population Census) records the population of each 2011 town in each decade going back to 1901, beginning with the first decade in which the Census defined a location to be urban.[29] Such an analysis is not possible for any other outcome, because urban population is the only variable available in a long panel spanning the many decades of canal construction.

To define whether a town is near a canal, we first draw a circle with a 20 km radius around each town. We define canal treatment for town i in year t as the percentage of the circle area that is overlapped by canal command areas. An alternate specification defines a binary treatment variable that takes the value 1 if more than 20% of the circle is covered by canal command areas.[30]

Equation 5.3 describes a standard two-way fixed effect (TWFE) continuous treatment difference-in-differences model to test whether town growth and emergence are affected by nearby canal construction:

$$y_{i,t} = \alpha_0 + \alpha_1 CANAL_{i,t} + \zeta_i + \nu_t + \epsilon_{i,t}. \tag{5.3}$$

Outcome $y_{i,t}$ is either an indicator for town existence, log(town population), or decadal growth in town i in decade t, and ζ_i and ν_t are town and decade fixed effects, respectively. When $y_{i,t}$

[29] We do not observe former towns which do not exist any longer, but given India's rising urbanization, town disappearance is very rare.

[30] Results are robust to alternate circle radius and threshold choices as shown in Appendix Table A16.

represents population, we assign the population value 2000 to towns that do not yet exist — this treats settlements before they become towns as if their size was just below the average population at which towns first appear in the data.[31] For the binary treatment, we use the estimator from Callaway and Sant'Anna (2020), using the not-yet-treated towns as the control group and defining treatment as the first decade when a town's 20 km radius catchment area is more than 20% covered by canal command areas. Standard errors are clustered at the subdistrict level.

6 Results

6.1 Direct Treatment Effects of Canals: Regression Discontinuity Estimates

We first report RDD estimates of the direct effects of canal access on irrigation outcomes, the mechanism through which we expect all other equilibrium effects to occur. Panel A of Table 3 shows that in canal-treated areas, 7.3 percentage points more of the land under cultivation is irrigated (17% more than in control settlements), and 9.5 percentage points (297%) more land is irrigated by canals. There are no discernible changes in other sources of irrigation. We test separately for effects on tubewell use, which would suggest greater groundwater access (for example, if canals recharge aquifers and find no effects in the RDD.

Panel B in Table 3 reports direct effects of canal access on agricultural outcomes. Canal-treated settlements experience higher agricultural productivity, with effects concentrated in the relatively dry winter (*rabi*) growing season. Treatment settlements have 7.3% higher values of satellite-derived land productivity in the dry season; productivity effects are positive but much smaller in the wet (*kharif*) season (1.8%, $p{=}0.051$). As expected, canals are more important during the dry winter growing season and only marginally affect productivity in the higher precipitation monsoon summer season. Settlements below canals also cultivate 2.7 percentage points more of their total land area, a 5% increase over control settlements, and are also 5% more likely to list a water-intensive crop (rice, cotton, or sugarcane) as one of their three primary crops. We find no evidence of increased capital intensity in agriculture, as measured by the share of households owning mechanized farm equipment.

[31] Of the 7,526 towns present in 2011, only 1,502 existed in 1911. We find similar results if we use 1 for the population of locations before they were urban, but we think that 2,000 is a better estimate of the population of pre-urban settlements.

The key question of this paper is how these major changes in agricultural productivity affect living standards and the growth of the non-farm economy. Panel C presents estimates of the impacts of canals on population density, non-farm employment, and predicted consumption. The only significant effect is on population: by 2011, treatment settlements have 16% more people per square kilometer than control settlements. This population gain could be the result of reduced out-migration, increased in-migration, increased fertility, or reduced mortality. We do not have data on migration flows or past mortality and fertility to distinguish these hypotheses directly. However, we can examine whether there are contemporary differences in fertility and mortality in canal villages, respectively proxied by the share of the population aged 0–6 and 70+. We find only a small 1.4% negative effect on the young population share, which goes in the opposite direction of the population gain (Appendix Table A1), and no effects on the 70+ population share.

Despite large gains in the productivity of the dominant economic sector in villages, we find no significant difference in living standards between above- and below-canal villages. The point estimate on log consumption is +0.008, with a 95% confidence interval of [−0.003,0.018]: we can rule out even small effects. There is also no evidence of structural transformation as measured by non-farm jobs per adult; nor do we find significant effects when we disaggregate employment into manufacturing, services, and agro-processing, the sector with the strongest linkage to agricultural production (Appendix Table A1). Total non-farm employment is higher than in canal settlements (as would be expected given the increase in population) but the non-farm *share* of the economy (the outcome of interest) is unchanged. Canal settlements have higher human capital (Panel D of Table 3); we measure a small but precise increase in the share of the adult population that has completed primary, middle, and secondary school, as well as the population literacy rate.

Figure 2 shows regression discontinuity binscatters of key outcomes in each of the categories above, with outcomes residualized on fixed effects and geophysical controls, showing the treatment effect at the RDD threshold, providing visual evidence of the effects of canals. Figure 3 plots the coefficients and 95% confidence intervals for the RDD coefficients reported in Table 3, normalized by the standard deviation of each variable in the control sample. The effect on population density is substantively larger than any other non-agricultural outcome.

The model in Section 3 suggests that the long-run spatial equilibrium will be characterized by equalization of returns to mobile factors (such as labor), but not to fixed factors (such as land). In the absence of high resolution data on wages and land rents, we proxy the returns to these factors by estimating canal treatment effects on predicted consumption separately for landless households (who own only labor) and for land-owning households (who own both land and labor).[32]

The results on land ownership are presented in Figure 4 and Table 4. Panel A of Table 4 shows a decline in the share of the population that are landowners in canal settlements relative to control settlements, with the average landholding size of landowners unchanged. This implies that the population increase in below-canal settlements is disproportionately driven by an increase in the number of landless households. The consumption effects of canals are substantially different for landed and landless households (Panel B of Table 4): there are no significant consumption effects for landless households, but landowner consumption is 2 percent higher in below-canal settlements; this result is statistically significantly different from the estimate for landless consumption at the 1% level. Partitioning landowners by nationally-defined land quartiles, effects increase monotonically by quartile, with no significant consumption effects on those owning <1 hectare of land (the 1st quartile), and a 2.7% effect on consumption for those in the top quartile owning >4 hectares (Panel B).[33] Both landless and landowning households experience gains in educational attainment, but effects for landowners are two to three times higher than for the landless (Table 4 Panel C). In short, the results are consistent with a model where the agricultural productivity gains from canals draw in new landless labor until a spatial equilibrium is reached, with equal landless wages in canal and non-canal areas, as we discuss further in Section 7.

6.1.1 Robustness

The RDD results are robust to alternate parameter choices. To show robustness, we replicate all of our primary outcomes in Appendix Tables A3 through A6; the different panels of the table show the result

[32]The predicted consumption measure is based on the ownership of a wide range of assets, so these proxies should be thought of as the real, rather than nominal, returns to labor and land.

[33]We define quartiles in the landholding distribution based on national data, to maintain consistent quartile boundaries across settlements. The first quartile owns 0-1 hectare of land, the second owns 1-2 hectares, the third owns 2-4 hectares, and the fourth owns more than 4 hectares.

of different specifications for each outcome. In Panel A of each table, we return to the sample with settlements in the donut hole (within 2.5 m of elevation of the canal) and those from subdistricts with unbalanced ruggedness. Panel B shows results excluding the donut hole but balanced on ruggedness. Panels C and D show results where settlement elevation is defined as the median pixel in the village and the 25th percentile pixel, respectively, rather than the 5$^{\text{th}}$ percentile used in the main analysis. To ensure that the variation is driven by arbitrary differences in elevation rather than potentially endogenous decisions about precise canal placement, Panel E excludes settlements intersected by canals and Panel F adds an additional control variable for distance from the settlement to the nearest canal. Panel G estimates canal effects using the alternative command area boundary RDD described in Section 5.1, where distance to the command area boundary is the running variable rather than relative elevation.

The results are highly consistent across all of the specifications; differences that appear in more than one specification are noted here. Two out of the six estimates show small substitution away from groundwater use in canal-irrigated areas; it is not surprising to find some substitution of this kind, but the magnitude is small relative to the increase in canal irrigation.[34] In the command area specification, we find higher *kharif* productivity effects than in the *rabi* season; in all other specifications, *rabi* effects are substantially higher. Some of the specifications show small increases in the use of mechanized farm equipment. The structural transformation measures are highly robust: we never estimate more than a 0.3 percentage point change in the non-farm employment share in any sector. The population change effects are highly significant for all specifications except the estimate using median settlement elevation to define treatment; as noted above, this is likely because many settlements whose median pixel is above a canal in fact have substantial area that is irrigable by the canal.

Finally, we test for sensitivity of outcomes to different RDD parameter choices. Appendix Table A7 shows that treatment effects are highly stable in magnitude and significance across bandwidths (Panel A), ruggedness balance restrictions (Panel B), and maximum distance to canal (Panel C).

[34]The implications of our findings are similar even if there is substitution away from groundwater — it would still imply a reduction in productivity and irrigation cost in canal areas.

6.2 Estimates of spillovers of canals to above-canal settlements

We next study spillover effects by comparing both above- and below-canal regions to more distant regions, matching on and controlling for geophysical features, as described in Section 5.2 and Equation 5.2. In each regression in Table 5, the omitted category is the distant settlement. The coefficient "below canal" is the difference between canal-treated settlements (as defined by relative elevation) and distant settlements, an alternate estimator of the direct effect of access to irrigation. The coefficient "above canal" is the difference between distant settlements and settlements that are close to the canal but above it in elevation and hence non-irrigated. The "above canal" coefficient is the coefficient of interest for studying spillovers. If canals affect the economy of *unirrigated* villages in the vicinity of the canal, this coefficient will be different from zero.

We find no substantial spillover effects in the agricultural outcomes. Above-canal settlements are marginally more likely to be irrigated by canals, but the effect is tiny and likely results from imprecision in our elevation measure — some parts of above-canal villages can in fact be irrigated by canal water due to topographical features not captured by our shortest distance measure. We calculate a statistically insignificant 0.5–1 percentage point increase in groundwater irrigation, which is slightly higher in above-canal areas, consistent with the marginal effect described in the robustness section above. In short, substitution into groundwater appears to be small to negligible. We do find statistically significant increases in water-intensive crops in above-canal areas, which may be due either to small increases in water access or the diffusion of these crops from nearby canal-irrigated settlements.

Spillovers to non-agricultural outcomes are minimal. The spillover point estimate on population density is positive but statistically insignificant and one tenth of the point estimate on the below-canal indicator. We estimate tight zeroes for spillovers on non-farm employment shares. Consumption spillovers are if anything negative for the landless ($p=0.08$). These results rule out the possibility that rural settlements just outside the canal irrigation zone are developing substantial non-farm sectors. We discuss our interpretation of these findings in light of all other results in Section 7 below.

6.3 Difference-in-Difference Estimates of the Effects of Canals on Urban Growth

The empirical strategies used thus far are best suited for measuring broad changes that occur across many settlements, and are either in the canal region or in direct proximity to it. But if canals caused changes primarily in a small number of urban areas with market linkages to canal areas, the estimates above might not have the precision to capture such a concentrated effect.

In this section, we use the long panel of town populations to test whether town growth responds to regional canal construction. Figure 5 provides suggestive evidence that towns grow more quickly after canals are built in their vicinity. For the set of towns that can be observed for 30 years before and after a canal is built that satisfies the binary treatment definition above, we plot log population of each town against the decade relative to the date of canal construction. There is a clear increase in the rate of population growth in the 30 years following canal construction relative to the 30 years preceding construction.

Table 6 shows difference-in-difference estimates from Equation 5.3 of the effect of canal construction on town growth. Odd-numbered columns show the binary treatment with the Callaway and Sant'Anna (2020) estimator, and even-numbered columns use the canonical difference-in-difference setup with a continuous treatment, which is the share of the town's 20 km radius catchment area that is in a canal command area.

Towns are more likely to appear (Columns 1 and 2) and are more likely to have larger populations (Columns 3 and 4) following nearby canal construction. Alternately, overall population growth is faster for canal-treated towns, with the continuous treatment estimate suggesting that a town entirely surrounded by a canal command area would grow 7.2% faster than a similar town not treated at all (Column 6). The number of post-treatment observations in the panel is too small to empirically distinguish a functional form for the time path of the population change. In other words, it is difficult to measure whether canals affect urban population growth in perpetuity, or whether they result in a level change in population which is arrived at over several decades. We therefore consider both of these possibilities when we discuss the magnitude of these estimates in the next section.

7 Discussion

The three components of the empirical analysis, together with the model, help us generate an understanding of how canals reshape the local economic geography. Canals create sharp spatial discontinuities in agricultural productivity. In irrigated villages, the return to land rises, growers shift to more water-intensive crops, and demand for labor rises. The increased demand likely puts upward pressure on wages in the short run, but in the long run that we observe, workers are attracted by these higher wages and the population grows until wages are again equalized across space. This migration may take the form of in-migration from outside the region, but it could also occur through reduced out-migration by workers or for marriage, lower mortality, or higher fertility. In the new equilibrium, canal areas are more densely populated, but the returns to labor are equalized with non-canal areas. The returns to land, the fixed factor, remain higher in irrigated areas even in the long run.

We see little evidence of small-scale industrialization or structural transformation in the rural areas where canals contribute directly to agricultural productivity. Structural transformation does occur, but new non-farm work opportunities are concentrated in cities. We think of these as production clusters, which can take advantage of agglomeration externalities and natural advantages that make them ideal locations for non-agricultural economic activity. The nature of the links between growing urban markets and greater agricultural productivity in canal areas is difficult to explore further in our empirical setting, where historical data is available only for town population, but the literature suggests a range of potential mechanisms. Bustos et al. (2020) show that landowners in Brazil invested land rents in urban areas that were connected by banking networks. Land rents could also be used to finance migration, another channel for urbanization and wealth accumulation among landowners (Clemens, 2014). Other linkages between greater agricultural output and non-farm industry are suggested by Johnston and Mellor (1961).

We can conduct a back-of-the-envelope calculation to get an approximation of the scale of the net population movement caused by canals. To conduct such an exercise, we make several simplifying assumptions. First, we assume that our estimates are driven entirely by net population movement rather than by fertility or mortality.[35] Second, we need to transform the urban treatment effects from Table 6 into static changes in present-day urban populations. Column 3 suggests a static treatment effect of 8.5%, *i.e.* that canal towns are 8.5% larger than they would be in the absence of canals. Column 5 shows a decadal growth effect of 4.3%; compounding this by four decades (the median treatment time in the sample data) would suggest that canals have made these towns 18% larger today. We therefore use 8.5–18% for the range of urban treatment effects. Third, we assume that these urban treatment effects apply to towns with populations less than 100,000; it seems doubtful that canals have equally large treatment effects in India's largest cities, which are widely diversified with many sources of growth.[36] Finally, to estimate rural population flows, we use the estimates from Table 5, multiplying the below-canal and above-canal treatment effects on population (18% and 2%, respectively) by the number of villages in below- and above-canal catchment areas.

Under these assumptions, our calculation suggests that India's canals have drawn an additional 5–9 million people to cities and towns in canal regions, and an additional 32 million people to rural canal regions. Canals have thus created substantial changes in India's economic geography, even if their long-run effects on the dispersion of wages have been limited.

Determining the extent of aggregate structural transformation driven by canals would require knowing the counterfactual sectors of these population gains, which we do not observe. However, we know that the vast majority of of urban migrants (85%) in 2001 came from rural areas; if people who changed locations because of canals had the same provenance, then India's canals would account for 3–5% of India's change in urbanization since 1951.

By studying the effects of irrigation at different geographic scales, our results unify some of the findings in the prior literature. Foster and Rosenzweig (2004a) find that villages most exposed to the Green Revolution shifted their production structure *away* from industry and toward agriculture, the Matsuyama (1992) open economy result. But their study is limited to villages; the industrialization

[35]This assumption serves only to simplify the exposition; the changes in economic geography are important whether driven by migration, fertility, or mortality.

[36]Indeed, Appendix Table A17 suggests that canals primarily affect the growth of smaller towns, by showing the diff-in-diff estimate separately for different quartiles of city size in the decade before canal construction. Including India's big cities in our calculation would raise the urban population change number below by about 1.5x.

that we measure is concentrated and occurs at some distance from the villages exposed to higher agricultural productivity. Bustos et al. (2016) found that the direction of structural change depended crucially on whether the technical change was labor-augmenting; the introduction of genetically-modified soy freed up labor to work in industry. Crucially, the units of observation in that paper are Brazilian municipalities, which have populations in the tens of thousands and incorporate the equivalents of Indian villages and towns. Our results suggest that urban towns may be the key focal point for structural change when it occurs.

A limitation of our analysis is that, with only limited data going back to the construction times of canals, measuring the aggregate effects of canals is difficult and beyond the scope of this paper. For example, if labor is sufficiently mobile, then canals could raise wages equally throughout the country, such that there would be no relationship between distance to canals and wages. In our model, we treat the agriculturally productive region as a small economy, but in India a very large share of land has been irrigated with canals. It is therefore plausible that the labor-sending regions have also experienced higher wages as a result. We therefore remain agnostic on the nature of aggregate spillovers.

Our results are relevant for policy even in the presence of some large-scale spillovers. Many development policies seek to foster non-farm work in rural areas, hoping to mitigate the pull of cities and create structural change in dispersed villages. Canals have caused large-scale population movements and substantially increased land productivity, but there is little evidence of structural change in treated rural areas. There are evidently important economic forces causing non-agricultural work to be concentrated in cities; policy will be most effective when it recognizes this reality.

Our results shed light on several strands of the literature beyond structural change. Several papers have suggested that increased labor demand in agriculture may deter human capital investment, particularly among the poor or landless (Foster and Rosenzweig, 2004b; Shah and Steinberg, 2017). In the context of canals, increased labor demand was met in the long run by net population growth, mitigating these potentially adverse effects, such that human capital increased among both the landed and the landless. This result recalls other scenarios where new economic opportunities resulted in higher educational investments (Jensen, 2012; Heath and Mobarak, 2015; Adukia et al., 2020). Foster and Rosenzweig (2004b) suggest a mechanism for the effects on education: demand for school investment among the wealthier land-rich could have resulted in more schools, which ultimately provided benefits to the landless as well.

8 Conclusion

India's canal systems provide an ideal testing ground for examining the geographic relationship between agricultural productivity improvements and structural transformation. In the long run, we find that spatial equilibrium was restored primarily through substantial changes in the size of the landless population. Decades after canals were built, there are few differences in living standards between landless workers in canal and non-canal settlements, and irrigated villages have similar non-farm activity to unirrigated villages. However, structural transformation has taken place, with towns emerging and growing disproportionately in canal regions.

The limitations of our work arise from the impossibility of measuring labor flows directly in our context; we observe higher population levels in canal areas, but the data do not tell us from where these people came. Mobile laborers who settled in canal-irrigated settlements, changes in fertility, or changes in exogamous marriage patterns could explain what we observe in equilibrium. Disentangling this economic history is beyond the scope of this paper but would be valuable in completing the picture.

Many shorter term studies have found that rising agricultural wages can deter or delay industrialization. Our study suggests that, in the long run, these effects may be tempered by changes in the labor supply. Naturally, it is difficult to compare different contexts in different times and places. Most of India's canals were built in or before the License Raj era, when manufacturing investments were significantly inhibited by the state, and firms could not rapidly respond to changes in labor demand, potentially enhancing the role of mobile labor. Whether modern agricultural shocks will be equally equilibrated by labor flows remains an important question for future research.

Mobile workers pose challenges for applied empirical researchers by violating assumptions of population stability across treatment and control groups. Yet hundreds of millions of Indians report living in places other than those of their birth, and tens of millions more migrated temporarily for work on an annual basis. Our study suggests that these large, mobile populations are a powerful economic force that can affect policy outcomes substantially.

References

Adhvaryu, Achyuta, A. V. Chari, and Siddharth Sharma, "Firing Costs and Flexibility: Evidence from Firms' Employment Responses to Shocks in India," *Review of Economics and Statistics*, 2013, *95.*

Adukia, Anjali, Sam Asher, and Paul Novosad, "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction," *American Economic Journal: Applied Economics*, 2020, *12* (1), 348–76.

Allcott, Hunt and Daniel Keniston, "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America," *The Review of Economic Studies*, 2018, *85* (2), 695–731.

Alvarez-Cuadrado, Francisco and Markus Poschke, "Structural Change out of Agriculture: Labor Push Versus Labor Pull," *American Economic Journal: Macroeconomics*, 2011, *3* (3), 127–58.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 2020, *110*, 797–823.

_ , Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform," *World Bank Economic Review*, 2021.

Athey, Susan and Guido W. Imbens, "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 2017, *31* (2), 3–32.

Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken, "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia," *American Economic Review*, 2021, *111* (12), 3827–71.

Blakeslee, David, Aaditya Dar, Ram Fishman, Samreen Malik, Heitor Pelegrina, and Karan Singh, "Irrigation and the Spatial Pattern of Structural Transformation in India," 2021. Working paper.

_ , Ram Fishman, and Veena Srinivasan, "Way Down in the Hole: Adaptation to Long-term Water Loss in Rural India," *American Economic Review*, 2021, *110*, 200–224.

Bose, Sugata, Bose Sugata et al., *Peasant Labour and Colonial Capital: Rural Bengal Since 1770*, Vol. 2, Cambridge University Press, 1993.

Burlig, Fiona and L. Preonas, "Out of the Darkness and Into the Light? Development Effects of Rural Electrification," 2021. Working paper.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural Productivity and Structural Transformation: Evidence from Brazil," *American Economic Review*, 2016, *106*, 1320–1365.

— **, Gabriel Garber, and Jacopo Ponticelli**, "Capital Accumulation and Structural Transformation," *The Quarterly Journal of Economics*, 2020, *135*, 1037–1094.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with Multiple Time Periods," *Journal of Econometrics*, 2020.

Clemens, Michael A, "Does Development Reduce Migration?," in "International handbook on migration and economic development," Edward Elgar Publishing, 2014.

Colmer, Jonathan, "Temperature, Labor Reallocation, and Industrial Production: Evidence from India," *American Economic Journal: Applied Economics*, 2021, *13*, 101–124.

Comin, Diego, Danial Lashkari, and Martí Mestieri, "Structural change with long-run income and price effects," *Econometrica*, 2021, *89* (1), 311–374.

Douie, James, "The Punjab Canal Colonies," *Journal of the Royal Society of Arts*, 1914, *62*, 611–623.

Elbers, Chris, Jean Lanjouw, and Peter Lanjouw, "Micro-level Estimation of Poverty and Inequality," *Econometrica*, 2003, *71*, 355–364.

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

Eswaran, Mukesh and Ashok Kotwal, "A Theory of Real Wage Growth in LDCs," *Journal of Development Economics*, 1993, *42* (2), 243–269.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 2014, *81*, 1046–1070.

Foster, Andrew D. and Mark R. Rosenzweig, "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review*, 1996, *86*, 931–953.

— **and** — , "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000," *Economic Development and Cultural Change*, 2004, *52*, 509–542.

— **and** — , "Technological Change and the Distribution of Schooling: Evidence from Green-Revolution India," *Journal of Development Economics*, 2004, *74*, 87–111.

— **and** — , "Chapter 47: Economic Development and the Decline of Agricultural Employment," *Handbook of Development Economics*, 2007, *4*, 3051–3083.

— **and** — , "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size," *NBER Working Paper*, 2017.

Funk, Chris C., Pete J. Peterson, Martin F. Landsfeld, Diego H. Pedreros, James P. Verdin, James D. Rowland, Bo E. Romero, Gregory J. Husak, Joel C. Michaelsen, and Andrew P. Verdin, "A quasi-global precipitation time series for drought monitoring," *Data Series*, 2014.

Funk, Chris, Pete Peterson, Seth Peterson, Shraddhanand Shukla, Frank Davenport, Joel Michaelsen, Kenneth R. Knapp, Martin Landsfeld, Gregory Husak, Laura Harrison, James Rowland, Michael Budde, Alex Meiburg, Tufa Dinku, Diego Pedreros, and Nicholas Mata, "A High-Resolution 1983-2016 Tmax Climate Data Record Based on Infrared Temperatures and Stations by the Climate Hazard Center," *Journal of Climate*, 9 2019, *32*, 5639–5658.

Gelman, Andrew and Guido Imbens, "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Gollin, Douglas, Casper Worm Hansen, and Asger Mose Wingender, "Two Blades of Grass: The Impact of the Green Revolution," *Journal of Political Economy*, 2021, *129*, 2344–2384.

_ , Stephen L. Parente, and Richard Rogerson, "The Food Problem and the Evolution of International Income Levels," *Journal of Monetary Economics*, 2007, *54*, 1230–1255.

_ , Stephen Parente, and Richard Rogerson, "The Role of Agriculture in Development," *American Economic Review*, 2002, *92*, 160–164.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya, "3G Internet and Confidence in Government," *The Quarterly Journal of Economics*, 2021, *136* (4), 2533–2613.

Hainmueller, Jens, "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political analysis*, 2012, *20* (1), 25–46.

Heath, Rachel and A Mushfiq Mobarak, "Manufacturing Growth and the Lives of Bangladeshi Women," *Journal of Development Economics*, 2015, *115*, 1–15.

Hornbeck, Richard and Pinar Keskin, "Does Agriculture Generate Local Economic Spillovers? Short-run and Long-run Evidence from the Ogallala Aquifer," *American Economic Journal: Economic Policy*, 2015, *7* (2), 192–213.

Iacus, Stefano M, Gary King, and Giuseppe Porro, "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 2012, *20* (1), 1–24.

Imbens, Guido W and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Imbert, Clément and John Papp, "Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India," *Journal of the European Economic Association*, 2020, *18* (2), 927–963.

Jain, Rajni, Prabhat Kishore, and Dhirendra Kumar Singh, "Irrigation in India: Status, Challenges and Options," *Journal of Soil and Water Conservation*, 2019, *18*, 354.

Jensen, Robert, "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India," *The Quarterly Journal of Economics*, 2012, *127* (2), 753–792.

Johnston, Bruce and John Mellor, "The Role of Agriculture in Economic Development," *American Economic Review*, 1961, *51*, 566–593.

Jones, Maria, Florence Kondylis, John Loeser, and Jeremy Magruder, "Factor Market Failures and the Adoption of Irrigation in Rwanda," *National Bureau of Economic Research*, 2020.

Jorgenson, Dale W, "The Development of a Dual Economy," *The Economic Journal*, 1961, *71* (282), 309–334.

Kone, Zovanga L, Maggie Y Liu, Aaditya Mattoo, Caglar Ozden, and Siddharth Sharma, "Internal Borders and Migration in India," *Journal of Economic Geography*, 2018, *18* (4), 729–759.

Kouadio, Louis, Nathaniel K. Newlands, Andrew Davidson, Yinsuo Zhang, and Aston Chipanshi, "Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale," *Remote Sensing*, 2014, *6*, 10193–10214.

Labus, M. P., G. A. Nielsen, R. L. Lawrence, R. Engel, and D. S. Long, "Wheat Yield Estimates Using Multi-temporal NDVI Satellite Imagery," *International Journal of Remote Sensing*, 2002, *23*, 4169–4180.

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

Matsuyama, Kiminori, "Agricultural Productivity, Comparative Advantage and Economic Growth," *Journal of Economic Theory*, 1992, *58*, 317–334.

Mellor, J. W., "Agriculture on the Road to Industrialization.," *Development Strategies Reconsidered*, 1986, pp. 67–89.

Mukherji, Aditi, "Evolution of Irrigation Sector," *Economic and Political Weekly*, 2016, *51.*

Ngai, Rachel and Christopher A. Pissarides, "Structural Change in a Multi-sector Model of Growth," *American Economic Review*, 2007, *97* (1), 429–443.

Nunn, Nathan and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Nurkse, Ragnar, "Some International Aspects of the Problem of Economic Development," *The American Economic Review*, 1952, *42* (2), 571–583.

Ranis, Gustav and John CH Fei, "A Theory of Economic Development," *American Economic Review*, 1961, *51* (4), 533–565.

Rasmussen, M. S., "Operational Yield Forecast Using AVHRR NDVI Data: Reduction of Environmental and Inter-annual Variability," *International Journal of Remote Sensing*, 1997, *18*, 1059–1077.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis," *Journal of Monetary Economics*, 2008, *55*, 234–250.

Riley, Shawn J., Stephen Daniel Degloria, and S.D. Elliot, "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 1999, *5*, 23–27.

Santangelo, Gabriella, "Firms and farms: The impact of agricultural productivity on the local Indian economy," 2019. Working paper.

Schultz, Theodore W, *The Economic Organization of Agriculture*, McGraw-Hill Book. Co, 1953.

_ , *Transforming Traditional Agriculture*, New Haven: Yale University Press, 1964.

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

Selvaraju, R., "Impact of El Niño-southern Oscillation on Indian Foodgrain Production," *International Journal of Climatology*, 2003, *23*, 187–206.

Shah, Manisha and Bryce Millett Steinberg, "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital," *Journal of Political Economy*, 2017, *125* (2), 527–561.

Shah, Tushaar, *Past, Present, and the Future of Canal Irrigation in India*, IDFC Institute, 2011.

Son, N. T., C. F. Chen, C. R. Chen, V. Q. Minh, and N. H. Trung, "A Comparative Analysis of Multitemporal MODIS EVI and NDVI Data for Large-Scale Rice Yield Estimation," *Agricultural and Forest Meteorology*, 2014, *197*, 52–64.

Timmer, C. Peter, "Chapter 8: The Agricultural Transformation," *Handbook of Development Economics*, 1988, *1*, 275–331.

Wardlow, Brian D. and Stephen L. Egbert, "A Comparison of MODIS 250-m EVI and NDVI Data For Crop Mapping: A Case Study for Southwest Kansas," *International Journal of Remote Sensing*, 2010, *31*, 805–830.

World Bank, *World Development Report 2008: Agriculture for Development*, The World Bank, 2007.

Table 1: Summary statistics

	All India	All canal-area settlements	All canal-area settlements minus donut hole	Ruggedness-balanced analysis sample
Sample Size	589,950	245,531	132,969	91,465
Percent Treatment	–	83	77	80
Means				
Total irrigated area (share of ag. land)	0.464	0.584	0.518	0.536
Canal irrigated area (share of ag. land)	0.134	0.177	0.138	0.132
Tubewell irrigated area (share of ag. land)	0.196	0.264	0.225	0.242
Other irrigated area (share of ag. land)	0.142	0.158	0.175	0.179
Agricultural land (share of total village area)	0.577	0.666	0.626	0.644
Kharif agricultural production, EVI-derived (log)	7.560	7.739	7.710	7.688
Rabi agricultural production, EVI-derived (log)	7.231	7.370	7.290	7.292
Any water intensive crop grown	0.586	0.657	0.604	0.605
Mechanized farming equipment (share of households)	0.047	0.062	0.055	0.061
Population density (log)	5.065	5.678	5.488	5.524
Consumption (log)	9.726	9.755	9.750	9.760
Total non-farm employment (share of adult pop)	0.096	0.085	0.088	0.085
Services employment (share of adult pop)	0.066	0.058	0.059	0.058
Manufacturing employment (share of adult pop)	0.019	0.020	0.020	0.020
Primary school ed attained (share of adult pop)	0.471	0.498	0.490	0.495
Middle school ed attained (share of adult pop)	0.318	0.339	0.329	0.331
Secondary school ed attained (share of adult pop)	0.194	0.212	0.207	0.207
Literacy rate (literate share of adult pop)	0.561	0.577	0.576	0.579

Notes: There are 589,950 settlements included in the All India sample, which is every village or town recorded in the 2011 Population Census with a non-zero population. In the second column, the All canal-area settlements sample includes towns and villages $\leq$10km from the nearest canal, and within 50m of the nearest canal in terms of elevation. In the third column, removing the donut hole drops settlements $\pm$2.5m in elevation from the nearest canal from the sample. We then impose a balance criteria on ruggedness by dropping settlements from subdistricts in which there is a $\geq$25% difference in average ruggedness between below-canal (treatment) and above-canal (control) settlements. The resulting sample, with 91,465 settlements, is the ruggedness-balanced analysis sample and is our preferred sample used in the RDD analysis. Note that the mean values reported for the ruggedness-balanced analysis sample also exclude subdistricts that do not contain at least one settlement in each of the treatment and control groups. All mean values are weighted by land area.

Table 2: Balance in the regression discontinuity design

	Ruggedness (TRI)	Annual rainfall avg. 2010-2014 (mm)	Max monthly temp. avg. 2010-2014 (°C)	
Below canal	0.055	-1.085	0.038***	
	(0.061)	(1.685)	(0.007)	
Control group mean	4.754	1167.092	32.163	
Observations	91,465	91,465	91,465	
R^2	0.620	0.988	0.981	
	Distance to coast (km)	**Distance to river (km)**	**Wetland rice (GAEZ)**	**Wheat (GAEZ)**
Below canal	-0.230	-2.097***	0.002	0.002
	(0.380)	(0.430)	(0.012)	(0.004)
Control group mean	362.339	24.564	2.272	0.767
Observations	91,465	91,465	91,465	91,465
R^2	0.999	0.876	0.923	0.983

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates for geophysical variables following Equation 5.1, dropping each outcome variable from the list of controls for each result. The Terrain Ruggedness Index (TRI) is a topographic measure of ruggedness, or how extreme elevation changes are in a given area, and was calculated following Riley et al. (1999) and Nunn and Puga (2012). Annual total rainfall was extracted from the Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) product produced by Funk et al. (2014). Average maximum monthly temperature was extracted from the Climate Hazards Center Infrared Temperature with Stations (CHIRTS) product released by Funk et al. (2019). Crop suitability measures are taken from the Global Agro-Ecological Zones (GAEZ) model that estimates expected conditions for agricultural production based on climate, soil, and terrain parameters. GAEZ model estimates made assuming gravity-fed irrigation and intermediate level inputs are used.

Table 3: Regression discontinuity results for main outcomes

Panel A. Irrigation outcomes

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Below canal	0.073***	0.095***	-0.005	-0.006
	(0.008)	(0.007)	(0.007)	(0.005)
Control group mean	0.430	0.032	0.210	0.195
Observations	83,182	83,192	83,247	82,385
R^2	0.61	0.38	0.47	0.63

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Below canal	0.027***	0.018*	0.073***	0.028***	0.002
	(0.005)	(0.009)	(0.012)	(0.009)	(0.002)
Control group mean	0.598	7.686	7.209	0.560	0.056
Observations	90,137	90,096	89,836	70,260	86,115
R^2	0.61	0.82	0.70	0.73	0.30

Panel C. Non-farm outcomes

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log)
Below canal	0.158***	0.000	0.002	-0.001	0.008
	(0.027)	(0.002)	(0.001)	(0.001)	(0.005)
Control group mean	5.241	0.089	0.058	0.020	9.744
Observations	91,465	85,342	85,342	85,342	86,842
R^2	0.42	0.25	0.19	0.28	0.53

Panel D. Education outcomes

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (literate share of pop.)
Below canal	0.012***	0.012***	0.009***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)
Control group mean	0.478	0.313	0.197	0.570
Observations	86,068	86,068	86,068	91,465
R^2	0.57	0.56	0.53	0.58

$^{*}p<0.10,^{**}p<0.05,^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for all outcomes variables. Each outcome variable is estimated separately, with the β_1 coefficient of the estimate reported in the first row with stars indicating its significance and the standard error below in parentheses. The control group mean (weighted by land area), the number of observations with non-missing data for the particular outcome variable, and the adjusted R^2 for each regression estimate are also shown.

Table 4: Regression discontinuity results for outcomes disaggregated by land ownership

Panel A. Land ownership overview

	Land-owning HHs (share of all HHs)	Avg. size of land holdings (log hectares, all HHs)	Avg. size of land holdings (log hectares, land-owning HHs)
Below canal	-0.025***	-0.054***	0.001
	(0.005)	(0.018)	(0.013)
Control group mean	0.536	0.750	1.526
Observations	86,117	83,796	83,763
R^2	0.460	0.460	0.500

Panel B. Consumption distribution

	Consumption pc		Consumption pc (log, land-owning HHs)			
	(log, landless HHs)	(log, land-owning HHs)	1st quartile	2nd quartile	3rd quartile	4th quartile
Below canal	0.002	0.020***	0.000	0.015**	0.019***	0.027***
	(0.006)	(0.005)	(0.008)	(0.006)	(0.006)	(0.006)
Control group mean	9.604	9.811	9.739	9.764	9.810	9.902
Observations	83,802	83,760	73,552	76,856	77,546	74,865
R^2	0.460	0.550	0.450	0.470	0.460	0.420

Panel C. Education attainment

	At least primary, share of		At least middle, share of		At least secondary, share of	
	landless pop.	land-owning pop.	landless pop.	land-owning pop.	landless pop.	land-owning pop.
Below canal	0.010***	0.021***	0.009***	0.021***	0.006**	0.018***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Control group mean	0.433	0.518	0.270	0.353	0.161	0.003
Observations	83,639	84,064	83,639	84,064	83,639	84,064
R^2	0.470	0.590	0.460	0.580	0.420	0.550

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for various outcomes pertaining to land ownership. Panel A shows estimates for the share of households that are landowners, the average size of land holdings for all households, and the average size of land holdings among land-owning households. Panel B first shows estimates for consumption disaggregated by land ownership status. Panel B then shows estimates for consumption by quartile, with each quartile of land-owning households separately estimated. The bottom (1^{st}) quartile are the land-owning households with total land holdings in the 0-25% range of the national distribution while the top (4^{th}) quartile are those in the top 75-100%. The quartile break points in ascending order are 1, 2, and 4 acres with each quartile excluding the bottom and including the top of the range defining the bin. Note that all consumption coefficients are in units of log consumption per capita, as they are throughout the paper.

Table 5: Comparison to distant settlements

Panel A. Irrigation outcomes

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Below canal	0.064***	0.085***	0.005	-0.016
	(0.015)	(0.010)	(0.008)	(0.010)
Above canal	0.010	0.007*	0.010	-0.007
	(0.008)	(0.004)	(0.007)	(0.005)
Control group mean	0.384	0.032	0.198	0.159
Observations	77,049	77,211	77,215	76,505
R^2	0.62	0.20	0.41	0.78

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Below canal	0.014*	0.009	0.022	0.047**	0.006**
	(0.008)	(0.015)	(0.019)	(0.023)	(0.003)
Above canal	-0.004	-0.004	-0.029	0.039**	0.000
	(0.008)	(0.011)	(0.019)	(0.018)	(0.002)
Control group mean	0.567	7.838	7.338	0.629	0.035
Observations	86,615	86,586	86,411	66,098	82,792
R^2	0.56	0.88	0.57	0.71	0.31

Panel C. Non-farm outcomes

	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Below canal	0.182***	0.002	0.003*	0.000	0.015*
	(0.032)	(0.003)	(0.001)	(0.002)	(0.008)
Above canal	0.020	0.004	0.002	0.001	-0.005
	(0.025)	(0.003)	(0.001)	(0.002)	(0.008)
Control group mean	5.504	0.086	0.053	0.021	9.626
Observations	87,756	80,546	80,546	80,546	83,262
R^2	0.27	0.14	0.09	0.22	0.43

Panel D. Outcomes disaggregated by land ownership

	Consumption pc (log, landless HHs)	Consumption pc (log) (log, land-owning HHs)	Middle school ed. (share of landless pop.)	Middle school ed. (share of land-owning pop.)
Below canal	-0.006	0.021**	0.009**	0.028***
	(0.007)	(0.009)	(0.004)	(0.007)
Above canal	-0.013*	0.004	0.002	0.008
	(0.007)	(0.009)	(0.004)	(0.005)
Control group mean	9.503	9.731	0.250	0.331
Observations	80,227	80,556	80,043	80,746
R^2	0.37	0.43	0.43	0.53

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for all outcomes variables. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal villages with respect to geophysical controls following Hainmueller (2012). The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

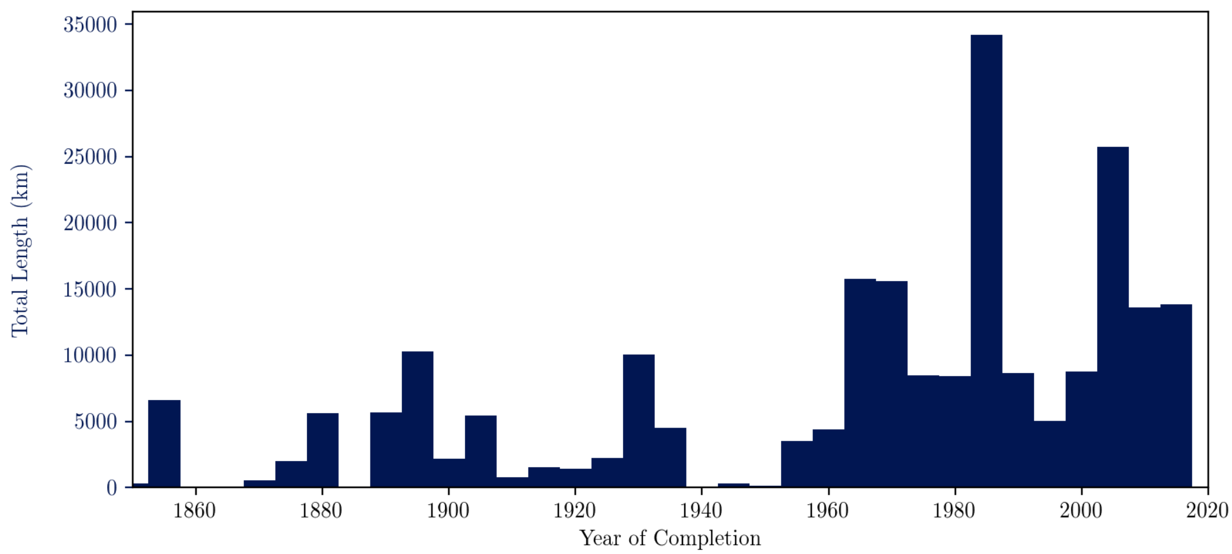
Table 6: Effect of canals on town growth

	Town Existence (pop. 5,000)		Population (log)		Growth (decadal)	
Command area in town catchment area	0.040***		0.085***		0.043**	
(binary treatment)	(0.014)		(0.026)		(0.020)	
Share of town catchment area in command area		0.090***		0.261***		0.072***
(continuous treatment)		(0.022)		(0.051)		(0.024)
Observations	21,636	46,932	21,636	46,932	19,833	43,021
R^2		0.67		0.82		0.06

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

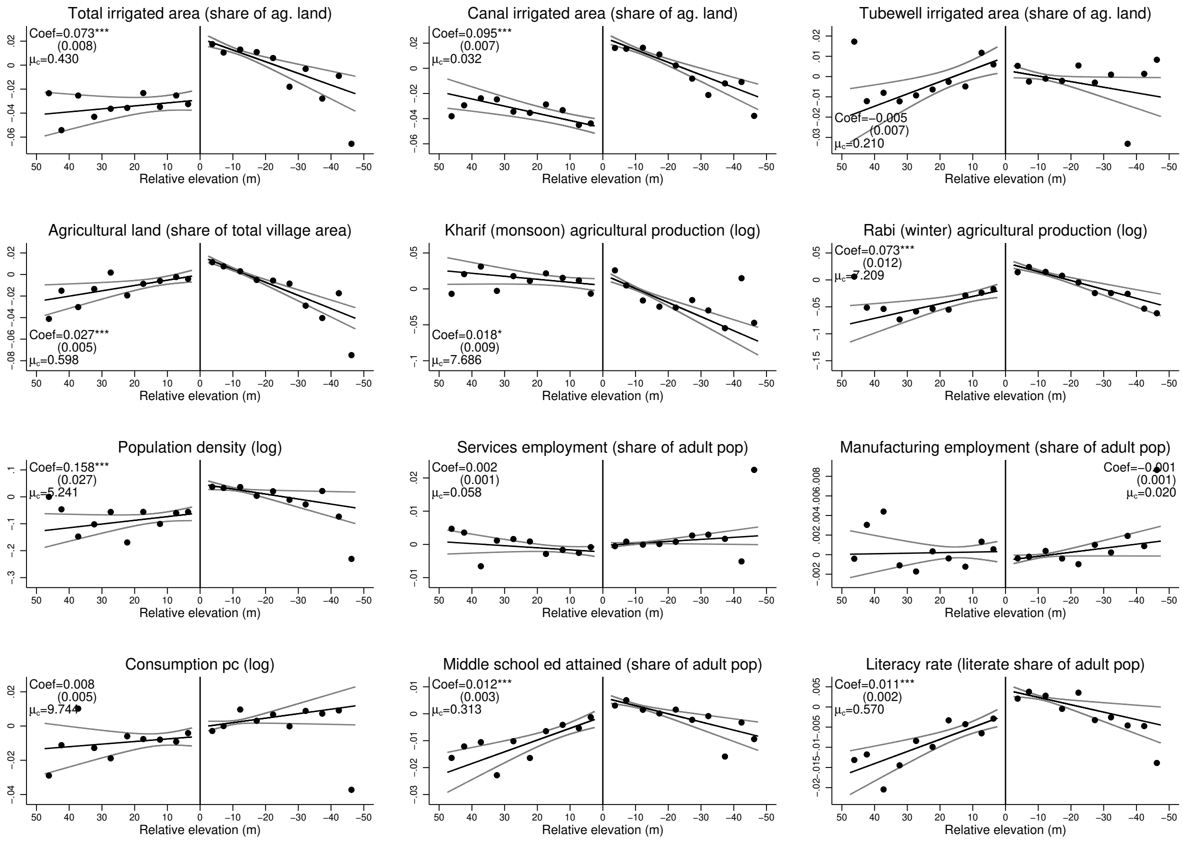
Notes: This table shows effect of canal construction on town growth as estimated by Equation 5.3. Each column reports the β_1 values for an independent estimate of various outcomes. The outcome variable in columns 1 and 2 is the existence of a town with population 5,000 or greater as the outcome variable. In columns 3 and 4 the outcome variable is log population, while in columns 5 and 6 it is decade-on-decade population growth. For all estimates, we assume that before a town appears in the time series, it is a settlement with a population of 2,000 (smaller than the population required to be declared a town). The first row shows results using a binary indicator for canal construction and a difference-in-difference estimator following Callaway and Sant'Anna (2020). For these estimates, a town is considered as treated when 20% of its catchment area (i.e., within a radius of 20km) has been covered by a command area. The second row shows results form a standard two-way fixed effect (TWFE) continuous treatment difference-in-differences, in which the share of the town catchment area covered by a command area is the dependent variable.

Figure 1: Canal construction through time



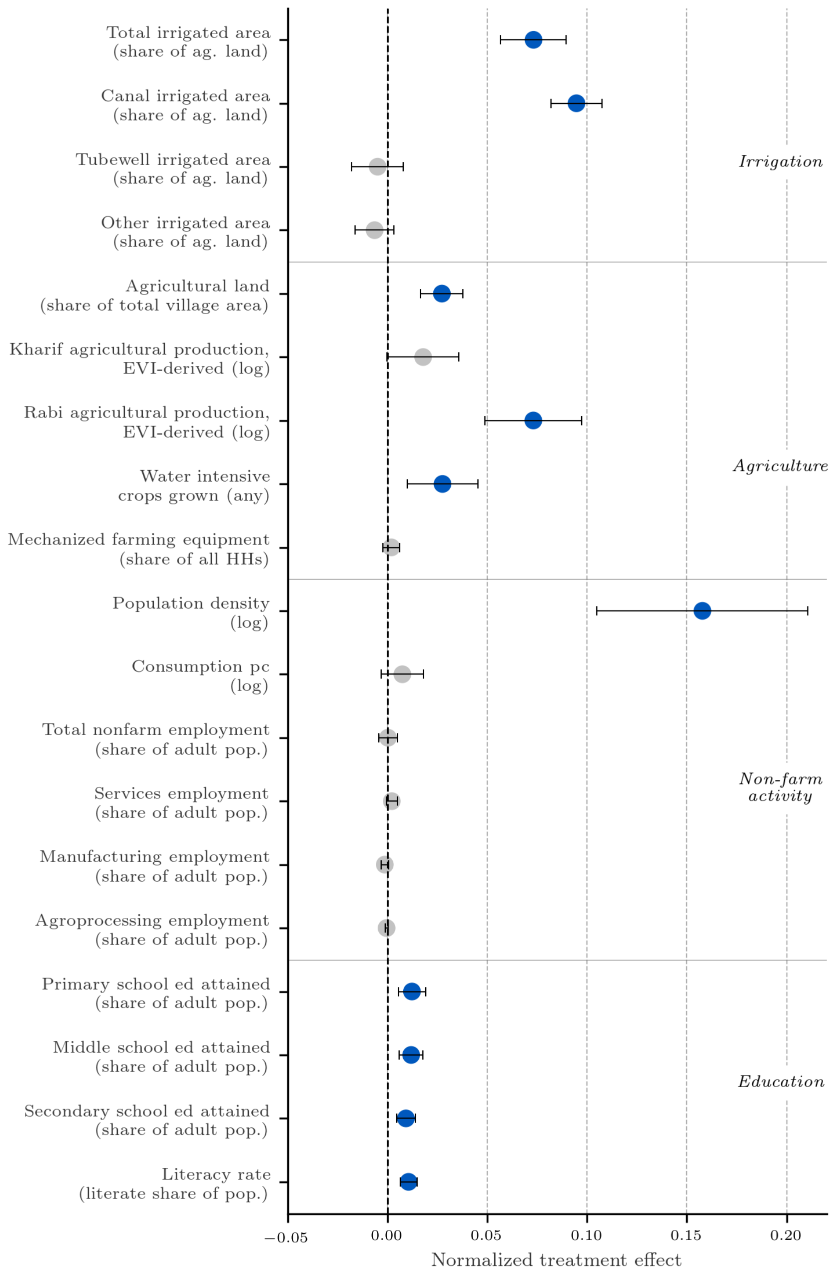
Notes: This plot shows the total length of medium and major canals constructed in India from 1850-2013. Any canals with dates older than 1850 are coded as 1850 while any canals not completed before 2013 are not included. Note that 217 of the 1442 total canal projects reported, or 9% of total canal length in the geospatial canals data, have an unknown date of completion and are not included in this plot. Additionally, 236 projects totaling 22% of total canal length in the data were not completed as of 2013 (the last date of our major outcomes) and so are not included in this plot.

Figure 2: Regression discontinuity binscatters for key outcomes



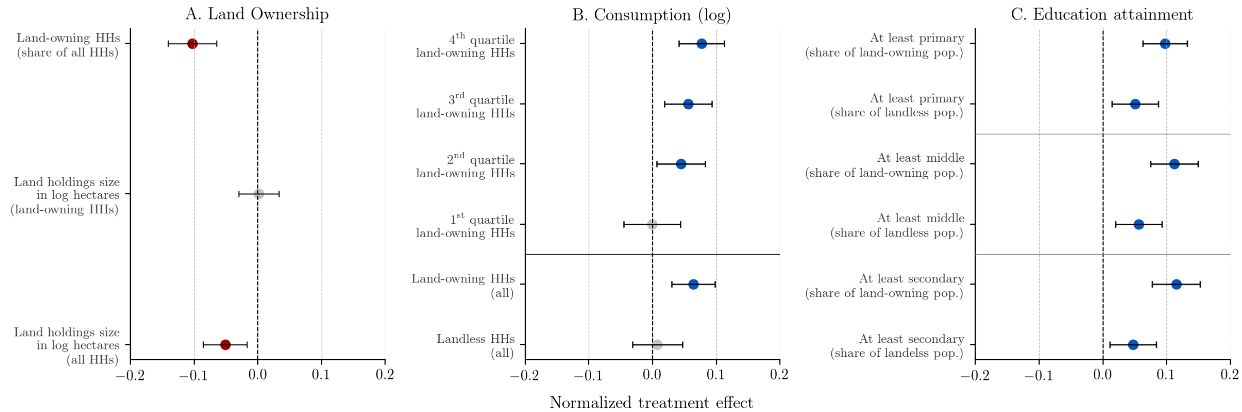
Notes: Each figure shows the binned scatterplot relationship between an outcome of interest and the RDD running variable (elevation relative to the nearest canal), after residualizing on the geophysical controls and subdistrict fixed effects. Below-canal (directly treated) settlements have negative relative elevation and lie to the right of the zero line, while above-canal (control) settlements have positive relative elevation and lie to the left of the zero line. All regressions follow Equation 5.1. The regression discontinuity coefficient (Coef) for each variable is reported with stars indicating the significance and the standard error in parentheses below. The control group mean is also reported (μ_c).

Figure 3: Regression discontinuity results for main outcomes



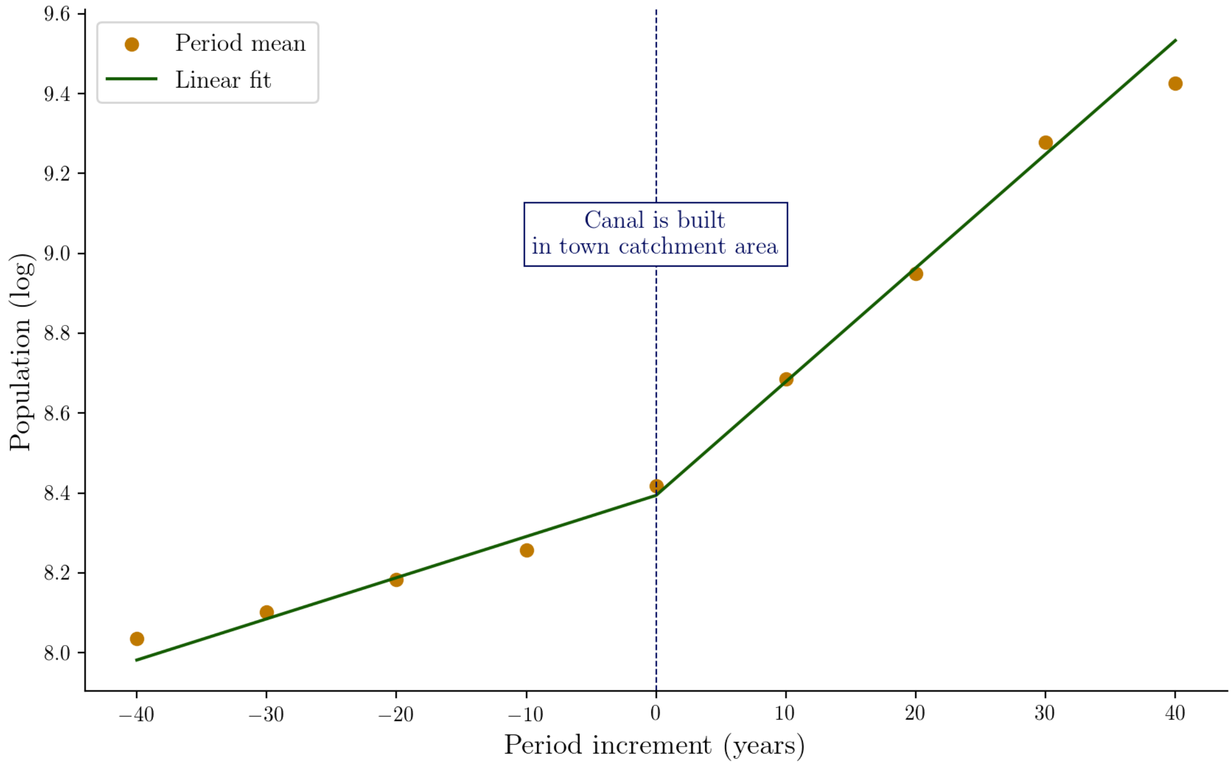
Notes: This figure shows the regression discontinuity estimates for the main outcomes variables following Equation 5.1 and reported in Table 3. Blue points indicate positive, significant normalized treatment effects while gray points indicate results not significant at the 95% level. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in control settlements of the analysis sample. Error bars indicate the 95% confidence interval for each estimate.

Figure 4: Land ownership outcomes



Notes: This figure shows the regression discontinuity estimates for various outcomes pertaining to land ownership, following Equation 5.1 and reported in Table 4. Blue points indicate positive, significant normalized treatment effects, red points indicate negative, significant normalized treatment effects, and gray points indicate results not significant at the 95% level. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in the control settlements of the analysis sample. Error bars indicate the 95% confidence interval for each estimate.

Figure 5: Trend break in town population growth after canal construction



Notes: This figure shows the trend break that occurs in town population after canal construction. Towns are aligned by period, where period 0 is the decade in which a command area first appeared in the town catchment area (defined by the circle of radius 20 km). Period increment 10 indicates 1 decade after the command area appearance, while period increment -10 indicates 1 decade before the first command area appearance.

A Appendix Tables and Figures

Table A1: Regression discontinuity results for additional outcomes

	Settlement is a town (likelihood)	Population age 0-6 (share of pop.)	Population age 70+ (share of pop.)	Agroprocessing emp. (share of adult pop.)
Below canal	0.010	-0.002***	0.000	0.000
	(0.007)	(0.001)	(0.000)	(0.000)
Control group mean	0.024	0.140	0.037	0.007
Observations	91,465	91,465	86,111	85,342
R^2	0.14	0.57	0.34	0.43

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for additional outcomes variables. Each outcome variable is estimated separately, with the β_1 coefficient of the estimate reported in the first row with stars indicating its significance and the standard error below in parentheses. The control group mean (weighted by land area), the number of observations with non-missing data for the particular outcome variable, and the adjusted R^2 for each regression estimate are also shown.

Table A2: Balance in the regression discontinuity design using distance to command area boundary

	Ruggedness (TRI)	Annual rainfall avg. 2010-2014 (mm)	Max monthly temp. avg. 2010-2014 (°C)	
Inside command area	-0.007	0.551	0.026***	
	(0.043)	(3.315)	(0.010)	
Control group mean	3.602	1317.632	31.976	
Observations	48,909	48,909	48,909	
R^2	0.640	0.990	0.990	
	Distance to coast (km)	Distance to river (km)	Wetland rice (GAEZ)	Wheat (GAEZ)
Outside command area	0.099	-0.701	-0.005	0.008
	(0.457)	(0.760)	(0.016)	(0.006)
Control group mean	467.522	23.429	2.704	1.188
Observations	48,909	48,909	48,909	48,909
R^2	1.000	0.930	0.960	0.990

$^{*}p<0.10,^{**}p<0.05,^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates for geophysical variables using the alternate command area boundary RDD, dropping each outcome variable from the list of controls for each result. This RDD specification uses distance to the command area boundary as the running variable instead of relative elevation. The Terrain Ruggedness Index (TRI) is a topographic measure of ruggedness, or how extreme elevation changes are in a given area, and was calculated following Riley et al. (1999) and Nunn and Puga (2012). Annual total rainfall was extracted from the Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) product produced by Funk et al. (2014). Average maximum monthly temperature was extracted from the Climate Hazards Center Infrared Temperature with Stations (CHIRTS) product released by Funk et al. (2019). Crop suitability measures are taken from the Global Agro-Ecological Zones (GAEZ) model that estimates expected conditions for agricultural production based on climate, soil, and terrain parameters. GAEZ model estimates made assuming gravity-fed irrigation and intermediate level inputs are used.

Table A3: Regression discontinuity results for irrigation outcomes (robustness specifications)

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Panel A: All canal-area settlements				
Below canal	0.063***	0.086***	-0.006	-0.009***
	(0.002)	(0.001)	(0.000)	(0.000)
Control group mean	0.435	0.070	0.214	0.158
Observations	230,114	230,165	230,257	228,287
R^2	0.650	0.450	0.470	0.560
Panel B: All canal-area settlements, minus donut hole				
Below canal	0.078***	0.106***	-0.006	-0.009**
	(0.002)	(0.002)	(0.000)	(0.000)
Control group mean	0.387	0.051	0.181	0.161
Observations	122,187	122,246	122,313	120,714
R^2	0.600	0.400	0.480	0.630
Panel C: Canal-area settlements balanced on ruggedness, using median settlement elevation				
Below canal	0.053***	0.071***	-0.007*	-0.005
	(0.002)	(0.004)	(0.000)	(-0.001)
Control group mean	0.469	0.068	0.245	0.162
Observations	90,418	90,419	90,480	89,476
R^2	0.640	0.500	0.500	0.620
Panel D: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation				
Below canal	0.073***	0.102***	-0.014**	-0.007*
	(0.002)	(0.003)	(0.000)	(0.000)
Control group mean	0.440	0.055	0.226	0.166
Observations	94,641	94,651	94,708	93,721
R^2	0.650	0.450	0.500	0.620
Panel E: Main analysis sample, excluding villages intersected by a canal				
Below canal	0.055***	0.043***	0.006	0.006
	(0.001)	(0.001)	(0.001)	(0.000)
Control group mean	0.410	0.048	0.204	0.165
Observations	60,803	60,787	60,826	60,361
R^2	0.660	0.350	0.510	0.670
Panel F: Main analysis sample, additional control for distance to canal				
Below canal	0.045***	0.047***	0.005	-0.001
	(0.001)	(0.000)	(0.001)	(0.000)
Control group mean	0.414	0.047	0.205	0.169
Observations	83,182	83,192	83,247	82,385
R^2	0.620	0.410	0.480	0.630
Panel G: Command area boundary RDD				
Inside command area	0.116***	0.154***	-0.008	-0.019*
	(-0.001)	(-0.002)	(-0.002)	(0.002)
Control group mean	0.535	0.063	0.333	0.156
Observations	43,244	43,206	43,239	42,752
R^2	0.730	0.510	0.570	0.520

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for irrigation outcomes using a range of samples to test the robustness of our results using our main analysis sample presented in Table 3. Panel A uses all settlements ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation. Panel B removes all settlements ±2.5m from the canal. The samples in Panel C and D employ the same sample definition as our main analysis sample, but instead of using the 5th percentile to parameterize settlement elevation, Panel C uses the median elevation while Panel D uses the 25th percentile. Panel E uses the main analysis sample but excludes settlements intersected by a canal branch, while Panel F keeps those settlements but adds an additional control for distance to the nearest canal. Panel G shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A4: Regression discontinuity results for agricultural outcomes (robustness specifications)

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water-intensive crops (any)	Mechanized farm equip. (share of all HHs)
Panel A: All canal-area settlements					
Below canal	0.030***	0.029***	0.052***	0.029***	0.003**
	(0.002)	(0.003)	(0.002)	(0.001)	(0.000)
Control group mean	0.594	7.744	7.272	0.698	0.043
Observations	242,138	242,195	240,398	193,838	231,547
R^2	0.590	0.790	0.720	0.730	0.340
Panel B: All canal-area settlements, minus donut hole					
Below canal	0.040***	0.027***	0.066***	0.028***	0.005**
	(0.002)	(0.003)	(0.002)	(0.001)	(0.000)
Control group mean	0.564	7.734	7.238	0.671	0.039
Observations	131,005	131,002	130,591	101,579	125,349
R^2	0.590	0.810	0.680	0.750	0.310
Panel C: Canal-area settlements balanced on ruggedness, using median settlement elevation					
Below canal	0.020***	0.025***	0.039***	0.012*	0.001
	(0.002)	(0.004)	(0.003)	(0.002)	(0.000)
Control group mean	0.629	7.747	7.292	0.691	0.049
Observations	97,491	97,424	97,061	77,302	93,397
R^2	0.610	0.820	0.730	0.740	0.350
Panel D: Canal-area settlements balanced on ruggedness, using 25$^{\text{th}}$ percentile settlement elevation					
Below canal	0.028***	0.034***	0.061***	0.023***	0.002
	(0.002)	(0.004)	(0.003)	(0.001)	(0.000)
Control group mean	0.612	7.737	7.273	0.667	0.048
Observations	101,914	101,871	101,590	80,405	97,498
R^2	0.620	0.820	0.730	0.720	0.340
Panel E: Main analysis sample, excluding villages intersected by a canal					
Below canal	0.019***	0.001	0.041***	0.023***	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Control group mean	0.594	7.745	7.248	0.648	0.046
Observations	66,670	66,656	66,513	52,385	63,457
R^2	0.620	0.830	0.730	0.750	0.320
Panel F: Main analysis sample, additional control for distance to canal					
Below canal	0.019***	-0.005	0.060***	0.016*	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Control group mean	0.593	7.736	7.244	0.649	0.047
Observations	90,137	90,096	89,836	70,260	86,115
R^2	0.610	0.830	0.710	0.730	0.310
Panel G: Command area boundary RDD					
Inside command area	0.025**	0.141***	0.059*	0.026	0.007**
	(0.000)	(-0.002)	(0.001)	(-0.002)	(0.000)
Control group mean	0.656	7.611	7.392	0.783	0.047
Observations	48,290	48,344	48,238	41,690	45,953
R^2	0.720	0.810	0.790	0.770	0.360

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for agricultural outcomes using a range of samples to test the robustness of our results using our main analysis sample presented in Table 3. Panel A uses all settlements $\leq$10km from the nearest canal in distance and $\pm$50m from the nearest canal in elevation. Panel B removes all settlements $\pm$2.5m from the canal. The samples in Panel C and D employ the same sample definition as our main analysis sample, but instead of using the 5$^{\text{th}}$ percentile to parameterize settlement elevation, Panel C uses the median elevation while Panel D uses the 25$^{\text{th}}$ percentile. Panel E uses the main analysis sample but excludes settlements intersected by a canal branch, while Panel F keeps those settlements but adds an additional control for distance to the nearest canal. Panel G shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A5: Regression discontinuity results for non-farm outcomes (robustness specifications)

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, landless HHs)	Consumption pc (log) (log, land-owning HHs)
Panel A: All canal-area settlements						
Below canal	0.127***	-0.002	0.001	-0.001*	0.003	0.018***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.676	0.095	0.060	0.019	9.535	9.736
Observations	245,506	226,890	226,890	226,890	225,297	224,745
R^2	0.480	0.260	0.200	0.240	0.470	0.590
Panel B: All canal-area settlements, minus donut hole						
Below canal	0.195***	0.002	0.003***	0.000	0.008	0.025***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.499	0.097	0.059	0.019	9.536	9.727
Observations	132,969	123,496	123,496	123,496	122,037	121,311
R^2	0.430	0.290	0.210	0.270	0.480	0.550
Panel C: Canal-area settlements balanced on ruggedness, using median settlement elevation						
Below canal	0.026	-0.003*	-0.003**	-0.001*	-0.001	0.008*
	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.796	0.090	0.057	0.019	9.542	9.757
Observations	99,086	92,479	92,479	92,479	91,015	90,847
R^2	0.460	0.310	0.230	0.290	0.500	0.590
Panel D: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation						
Below canal	0.119***	0.000	0.001	-0.001	0.004	0.019***
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.685	0.091	0.058	0.019	9.550	9.757
Observations	103,543	96,664	96,664	96,664	94,961	94,883
R^2	0.470	0.300	0.230	0.280	0.480	0.590
Panel E: Main analysis sample, excluding villages intersected by a canal						
Below canal	0.111***	0.000	0.002	-0.002**	0.007	0.011*
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	5.568	0.091	0.057	0.018	9.543	9.743
Observations	67,546	62,369	62,369	62,369	61,478	61,492
R^2	0.450	0.340	0.210	0.300	0.460	0.550
Panel F: Main analysis sample, additional control for distance to canal						
Below canal	0.094***	-0.002	0.001	-0.002**	-0.001	0.011**
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	5.570	0.090	0.056	0.019	9.550	9.748
Observations	91,465	85,342	85,342	85,342	83,802	83,760
R^2	0.430	0.260	0.200	0.290	0.470	0.560
Panel G: Command area boundary RDD						
Inside command area	0.254***	0.002	0.001	0.001	-0.001	0.035***
	(0.002)	(0.000)	(0.000)	(0.000)	(-0.001)	(0.000)
Control group mean	6.308	0.085	0.058	0.020	9.516	9.767
Observations	48,909	45,098	45,098	45,098	45,029	44,657
R^2	0.620	0.350	0.260	0.360	0.500	0.580

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for non-farm outcomes using a range of samples to test the robustness of our results using our main analysis sample presented in Table 3. Panel A uses all settlements ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation. Panel B removes all settlements ±2.5m from the canal. The samples in Panel C and D employ the same sample definition as our main analysis sample, but instead of using the 5th percentile to parameterize settlement elevation, Panel C uses the median elevation while Panel D uses the 25th percentile. Panel E uses the main analysis sample but excludes settlements intersected by a canal branch, while Panel F keeps those settlements but adds an additional control for distance to the nearest canal. Panel G shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A6: Regression discontinuity results for education outcomes (robustness specifications)

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (literate share of pop.)
Panel A: All canal-area settlement				
Below canal	0.014***	0.014***	0.011***	0.009***
	(0.001)	(0.001)	(0.000)	(0.000)
Control group mean	0.462	0.308	0.185	0.554
Observations	231,428	231,428	231,428	245,506
R^2	0.580	0.560	0.530	0.600
Panel B: All canal-area settlements, minus donut hole				
Below canal	0.020***	0.019***	0.015***	0.013***
	(0.001)	(0.001)	(0.000)	(0.000)
Control group mean	0.447	0.294	0.176	0.545
Observations	125,287	125,287	125,287	132,969
R^2	0.590	0.580	0.550	0.600
Panel C: Canal-area settlements balanced on ruggedness, using median settlement elevation				
Below canal	0.007***	0.005**	0.003	0.003*
	(0.001)	(0.001)	(0.001)	(0.001)
Control group mean	0.482	0.324	0.196	0.567
Observations	93,359	93,359	93,359	99,086
R^2	0.590	0.570	0.540	0.610
Panel D: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation				
Below canal	0.012***	0.011***	0.009***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Control group mean	0.475	0.317	0.192	0.564
Observations	97,455	97,455	97,455	103,543
R^2	0.580	0.570	0.550	0.610
Panel E: Main analysis sample, excluding villages intersected by a canal				
Below canal	0.008**	0.007**	0.006**	0.008***
	(0.001)	(0.001)	(0.000)	(0.001)
Control group mean	0.461	0.305	0.185	0.556
Observations	63,419	63,419	63,419	67,546
R^2	0.580	0.560	0.530	0.590
Panel F: Main analysis sample, additional control for distance to canal				
Below canal	0.005	0.006**	0.005**	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	0.465	0.308	0.187	0.558
Observations	86,068	86,068	86,068	91,465
R^2	0.580	0.560	0.540	0.590
Panel G: Command area boundary RDD				
Inside command area	0.021***	0.021***	0.020***	0.019***
	(-0.001)	(-0.001)	(0.000)	(0.000)
Control group mean	0.471	0.322	0.186	0.562
Observations	45,941	45,941	45,941	48,909
R^2	0.660	0.620	0.570	0.690

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for education outcomes using a range of samples to test the robustness of our results using our main analysis sample presented in Table 3. Panel A uses all settlements ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation. Panel B removes all settlements ±2.5m from the canal. The samples in Panel C and D employ the same sample definition as our main analysis sample, but instead of using the 5th percentile to parameterize settlement elevation, Panel C uses the median elevation while Panel D uses the 25th percentile. Panel E uses the main analysis sample but excludes settlements intersected by a canal branch, while Panel F keeps those settlements but adds an additional control for distance to the nearest canal. Panel G shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A7: Regression discontinuity results for primary outcomes (sensitivity analysis)

Panel A. Regression discontinuity bandwidth

Bandwidth (m)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (literate share of pop.)	Ruggedness (TRI)	Sample size
25	0.065***	0.054***	0.104***	0.009***	0.011	88,535
	(0.009)	(0.012)	(0.024)	(0.002)	(0.055)	
50	0.068***	0.080***	0.148***	0.011***	0.055	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.061)	
75	0.073***	0.072***	0.163***	0.011***	0.028	90,735
	(0.008)	(0.012)	(0.023)	(0.002)	(0.068)	

Panel B. Percent difference in ruggedness

Percent difference in ruggedness (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
10%	0.064***	0.046***	0.128***	0.006**	0.017	54,914
	(0.011)	(0.016)	(0.028)	(0.003)	(0.037)	
25%	0.073***	0.073***	0.148***	0.011***	0.055	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.061)	
50%	0.074***	0.068***	0.165***	0.011***	-0.073	116,695
	(0.007)	(0.011)	(0.019)	(0.002)	(0.055)	

Panel C. Distance to Canal

Max distance to canal (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
5	0.081***	0.061***	0.179***	0.008***	0.041	61,217
	(0.011)	(0.015)	(0.027)	(0.003)	(0.054)	
10	0.073***	0.073***	0.148***	0.011***	0.055	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.061)	
15	0.069***	0.078***	0.141***	0.003	0.018	109,071
	(0.007)	(0.011)	(0.020)	(0.007)	(0.047)	

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table shows the sensitivity of our regression discontinuity estimated following Equation 5.1 to changes in the construction of our sample. We show results for four primary outcomes and also for ruggedness, to test for balance in our primary geographic fundamental variable as the sample changes. Each outcome variable is estimated separately after one assumption has been changed to define the sample, with the β_1 coefficient of the estimate reported in the top row with stars indicating its significance and the standard error below in parentheses. The bolded parameters in each panel indicate the values uses in our main analysis sample. These preferred values are used for the two parameters not being tested in each panel. In Panel A, we modify the bandwidth of the regression discontinuity, where 50m would include settlements that lie 50m above to 50m below the nearest canal. Here we test 25m and 75m bandwidths in addition to our preferred 50m bandwidth. In Panel B, we modify the threshold allowed for the average difference in ruggedness between below- and above-canal settlements in a subdistrict. We test 10% (more strict) and 50% (less strict) in addition to our preferred 25% threshold. Lastly, in Panel C we modify the maximum distance a settlement may lie away from the nearest canal to be considered treated by that canal. Here we test 5km and 15km in addition to our preferred 10km.

Table A8: Comparison to distant settlements for irrigation outcomes (entropy balance robustness specifications)

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Panel A. Entropy balance, no outliers dropped				
Below canal	0.065***	0.089***	0.005	-0.018**
	(0.015)	(0.010)	(0.007)	(0.009)
Above canal	0.011	0.005	0.010	-0.005
	(0.008)	(0.003)	(0.007)	(0.005)
Control group mean	0.383	0.037	0.189	0.162
Observations	104,034	104,218	104,215	103,420
R^2	0.60	0.18	0.40	0.76
Panel B. Entropy balance, 1% outliers dropped				
Below canal	0.064***	0.090***	0.005	-0.019**
	(0.016)	(0.010)	(0.007)	(0.010)
Above canal	0.014*	0.007*	0.013*	-0.006
	(0.008)	(0.004)	(0.007)	(0.005)
Control group mean	0.384	0.035	0.195	0.158
Observations	91,761	91,938	91,940	91,173
R^2	0.60	0.17	0.40	0.76
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**				
Below canal	0.064***	0.085***	0.005	-0.016
	(0.015)	(0.010)	(0.008)	(0.010)
Above canal	0.010	0.007*	0.010	-0.007
	(0.008)	(0.004)	(0.007)	(0.005)
Control group mean	0.384	0.032	0.198	0.159
Observations	77,049	77,211	77,215	76,505
R^2	0.62	0.20	0.41	0.78
Panel D. Entropy balance, 5% outliers dropped				
Below canal	0.056***	0.091***	0.003	-0.028**
	(0.017)	(0.012)	(0.009)	(0.011)
Above canal	0.005	0.005	0.006	-0.007
	(0.009)	(0.005)	(0.008)	(0.007)
Control group mean	0.374	0.027	0.202	0.149
Observations	58,745	58,898	58,894	58,269
R^2	0.65	0.17	0.44	0.79

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for irrigation outcomes using a range of samples to test the robustness of our main results presented in Table 5. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal villages with respect to geophysical controls following Hainmueller (2012). Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main results in Table 5), and Panel D drops 5% outliers. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A9: Comparison to distant settlements for agricultural outcomes (entropy balance robustness specifications)

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water-intensive crops (any)	Mechanized farm equip. (share of all HHs)
Panel A. Entropy balance, no outliers dropped					
Below canal	0.015*	-0.002	0.040**	0.066***	0.001
	(0.009)	(0.019)	(0.020)	(0.022)	(0.003)
Above canal	-0.004	0.004	-0.022	0.038**	-0.001
	(0.009)	(0.015)	(0.017)	(0.017)	(0.003)
Control group mean	0.562	7.760	7.329	0.648	0.038
Observations	115,189	115,385	115,130	88,596	110,612
R^2	0.54	0.84	0.58	0.65	0.30
Panel B. Entropy balance, 1% outliers dropped					
Below canal	0.018**	0.004	0.036*	0.068***	0.004
	(0.008)	(0.019)	(0.019)	(0.022)	(0.002)
Above canal	0.001	0.001	-0.019	0.050***	0.001
	(0.008)	(0.014)	(0.018)	(0.017)	(0.002)
Control group mean	0.564	7.798	7.328	0.637	0.037
Observations	102,304	102,306	102,075	78,513	98,014
R^2	0.55	0.85	0.57	0.68	0.31
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**					
Below canal	0.014*	0.009	0.022	0.047**	0.006**
	(0.008)	(0.015)	(0.019)	(0.023)	(0.003)
Above canal	-0.004	-0.004	-0.029	0.039**	0.000
	(0.008)	(0.011)	(0.019)	(0.018)	(0.002)
Control group mean	0.567	7.838	7.338	0.629	0.035
Observations	86,615	86,586	86,411	66,098	82,792
R^2	0.56	0.88	0.57	0.71	0.31
Panel D. Entropy balance, 5% outliers dropped					
Below canal	0.016*	0.016	-0.006	0.039	0.004
	(0.009)	(0.015)	(0.023)	(0.028)	(0.003)
Above canal	-0.004	-0.003	-0.033	0.046*	0.003
	(0.008)	(0.011)	(0.023)	(0.027)	0.002
Control group mean	0.585	7.876	7.378	0.602	0.035
Observations	66,083	66,064	65,953	49,163	62,952
R^2	0.57	0.89	0.59	0.72	0.30

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for agricultural outcomes using a range of samples to test the robustness of our main results presented in Table 5. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal villages with respect to geophysical controls following Hainmueller (2012). Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main results in Table 5), and Panel D drops 5% outliers. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A10: Comparison to distant settlements for non-farm outcomes (entropy balance robustness specifications)

	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Panel A. Entropy balance, no outliers dropped					
Below canal	0.186***	0.003	0.003**	0.000	0.011
	(0.039)	(0.003)	(0.001)	(0.001)	(0.008)
Above canal	0.004	0.000	0.000	0.000	-0.013
	(0.033)	(0.003)	(0.001)	(0.002)	(0.009)
Control group mean	5.507	0.092	0.057	0.021	9.634
Observations	117,006	107,341	107,341	107,341	111,329
R^2	0.31	0.13	0.10	0.19	0.41
Panel B. Entropy balance, 1% outliers dropped					
Below canal	0.203***	0.003	0.003***	0.001	0.014*
	(0.030)	(0.003)	(0.001)	(0.001)	(0.008)
Above canal	0.030	0.003	0.001	0.001	-0.008
	(0.024)	(0.003)	(0.001)	(0.002)	(0.007)
Control group mean	5.512	0.088	0.055	0.021	9.632
Observations	103,722	95,296	95,296	95,296	98,625
R^2	0.27	0.13	0.09	0.20	0.42
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**					
Below canal	0.182***	0.002	0.003*	0.000	0.015*
	(0.032)	(0.003)	(0.001)	(0.002)	(0.008)
Above canal	0.020	0.004	0.002	0.001	-0.005
	(0.025)	(0.003)	(0.001)	(0.002)	(0.008)
Control group mean	5.504	0.086	0.053	0.021	9.626
Observations	87,756	80,546	80,546	80,546	83,262
R^2	0.27	0.14	0.09	0.22	0.43
Panel D. Entropy balance, 5% outliers dropped					
Below canal	0.191***	0.005	0.004***	0.002	0.021**
	(0.038)	(0.003)	(0.002)	(0.002)	(0.010)
Above canal	0.013	0.006	0.003	0.002	-0.004
	(0.027)	(0.004)	(0.002)	(0.002)	(0.008)
Control group mean	5.472	0.081	0.050	0.020	9.619
Observations	66,924	61,313	61,313	61,313	63,279
R^2	0.26	0.16	0.09	0.26	0.38

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for non-farm outcomes using a range of samples to test the robustness of our main results presented in Table 5. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal villages with respect to geophysical controls following Hainmueller (2012). Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main results in Table 5), and Panel D drops 5% outliers. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A11: Comparison to distant settlements for outcomes disaggregated by land ownership (entropy balance robustness specifications)

	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Panel A. Entropy balance, no outliers dropped					
Below canal	0.186***	0.003	0.003**	0.000	0.011
	(0.039)	(0.003)	(0.001)	(0.001)	(0.008)
Above canal	0.004	0.000	0.000	0.000	-0.013
	(0.033)	(0.003)	(0.001)	(0.002)	(0.009)
Control group mean	5.507	0.092	0.057	0.021	9.634
Observations	117,006	107,341	107,341	107,341	111,329
R^2	0.31	0.13	0.10	0.19	0.41
Panel B. Entropy balance, 1% outliers dropped					
Below canal	0.203***	0.003	0.003***	0.001	0.014*
	(0.030)	(0.003)	(0.001)	(0.001)	(0.008)
Above canal	0.030	0.003	0.001	0.001	-0.008
	(0.024)	(0.003)	(0.001)	(0.002)	(0.007)
Control group mean	5.512	0.088	0.055	0.021	9.632
Observations	103,722	95,296	95,296	95,296	98,625
R^2	0.27	0.13	0.09	0.20	0.42
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**					
Below canal	0.182***	0.002	0.003*	0.000	0.015*
	(0.032)	(0.003)	(0.001)	(0.002)	(0.008)
Above canal	0.020	0.004	0.002	0.001	-0.005
	(0.025)	(0.003)	(0.001)	(0.002)	(0.008)
Control group mean	5.504	0.086	0.053	0.021	9.626
Observations	87,756	80,546	80,546	80,546	83,262
R^2	0.27	0.14	0.09	0.22	0.43
Panel D. Entropy balance, 5% outliers dropped					
Below canal	0.191***	0.005	0.004***	0.002	0.021**
	(0.038)	(0.003)	(0.002)	(0.002)	(0.010)
Above canal	0.013	0.006	0.003	0.002	-0.004
	(0.027)	(0.004)	(0.002)	(0.002)	(0.008)
Control group mean	5.472	0.081	0.050	0.020	9.619
Observations	66,924	61,313	61,313	61,313	63,279
R^2	0.26	0.16	0.09	0.26	0.38

${}^{*}p<0.10, {}^{**}p<0.05, {}^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for various outcomes disaggregated by landownership using a range of samples to test the robustness of our main results presented in Table 5. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal villages with respect to geophysical controls following Hainmueller (2012). Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main results in Table 5), and Panel D drops 5% outliers. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A12: Comparison to distant settlements for irrigation outcomes (coarsened exact matching robustness specifications)

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Panel A. Coarsened exact matching, no outliers dropped				
Below canal	0.040***	0.064***	-0.005	-0.015*
	(0.012)	(0.008)	(0.009)	(0.009)
Above canal	-0.012	0.002	-0.001	-0.013*
	(0.010)	(0.006)	(0.008)	(0.007)
Control group mean	0.405	0.031	0.207	0.173
Observations	38,369	38,432	38,432	38,114
R^2	0.57	0.21	0.42	0.68
Panel B. Coarsened exact matching, 1% outliers dropped				
Below canal	0.039**	0.069***	-0.016	-0.011
	(0.016)	(0.010)	(0.013)	(0.012)
Above canal	-0.014	0.010*	-0.013	-0.011
	(0.012)	(0.006)	(0.012)	(0.010)
Control group mean	0.453	0.022	0.257	0.180
Observations	21,220	21,250	21,251	21,071
R^2	0.58	0.18	0.44	0.69
Panel C. Coarsened exact matching, 2.5% outliers dropped				
Below canal	0.020	0.062***	-0.018	-0.020
	(0.017)	(0.011)	(0.012)	(0.013)
Above canal	-0.018	0.005	-0.004	-0.016
	(0.015)	(0.008)	(0.012)	(0.011)
Control group mean	0.407	0.027	0.235	0.157
Observations	13,905	13,918	13,916	13,802
R^2	0.59	0.22	0.45	0.68
Panel D. Coarsened exact matching, 5% outliers dropped				
Below canal	0.016	0.057***	-0.011	-0.024
	(0.020)	(0.010)	(0.013)	(0.016)
Above canal	-0.009	0.004	0.003	-0.017
	(0.018)	(0.008)	(0.013)	(0.014)
Control group mean	0.421	0.018	0.235	0.174
Observations	8,078	8,087	8,089	8,019
R^2	0.60	0.20	0.47	0.69

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for irrigation outcomes using a range of samples to test the robustness of our main results presented in Table 5. Here we employ coarsened exact matching (CEM) following Iacus et al. (2012), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions, instead of entropy balancing to ensure the distant settlements are comparable to the above-canal settlements. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A13: Comparison to distant settlements for agricultural outcomes (coarsened exact matching robustness specifications)

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water-intensive crops (any)	Mechanized farm equip. (share of all HHs)
Panel A. Coarsened exact matching, no outliers dropped					
Below canal	0.010	-0.040***	0.025	0.021	0.009**
	(0.010)	(0.013)	(0.024)	(0.015)	(0.003)
Above canal	0.003	-0.002	0.008	0.010	0.004**
	(0.009)	(0.010)	(0.021)	(0.74)	(0.002)
Control group mean	0.622	7.876	7.346	0.578	0.041
Observations	43,666	43,642	43,530	33,568	41,819
R^2	0.58	0.75	0.61	0.74	0.24
Panel B. Coarsened exact matching, 1% outliers dropped					
Below canal	0.016	-0.025	0.016	0.001	0.007*
	(0.014)	(0.017)	(0.025)	(0.015)	(0.004)
Above canal	0.013	0.007	0.004	-0.003	0.004
	(0.012)	(0.015)	(0.023)	(0.014)	(0.003)
Control group mean	0.669	7.933	7.533	0.496	0.049
Observations	24,714	24,693	24,641	19,201	23,680
R^2	0.62	0.77	0.64	0.74	0.26
Panel C. Coarsened exact matching, 2.5% outliers dropped					
Below canal	0.001	-0.031*	0.033	0.062***	0.004
	(0.010)	(0.016)	(0.034)	(0.023)	(0.003)
Above canal	0.000	-0.005	0.026	0.047**	0.004
	(0.010)	(0.015)	(0.030)	(0.022)	(0.003)
Control group mean	0.665	7.908	7.416	0.562	0.054
Observations	16,371	16,372	16,331	12,768	15,646
R^2	0.64	0.78	0.67	0.73	0.26
Panel D. Coarsened exact matching, 5% outliers dropped					
Below canal	0.005	-0.027	0.045	0.006	-0.004
	(0.012)	(0.021)	(0.035)	(0.019)	(0.005)
Above canal	0.006	0.002	0.055**	-0.003	-0.003
	(0.012)	(0.019)	(0.027)	(0.017)	(0.005)
Control group mean	0.663	7.916	7.460	0.561	0.054
Observations	9,203	9,208	9,187	7,126	8,851
R^2	0.67	0.78	0.68	0.73	0.22

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for agricultural outcomes using a range of samples to test the robustness of our main results presented in Table 5. Here we employ coarsened exact matching (CEM) following Iacus et al. (2012), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions, instead of entropy balancing to ensure the distant settlements are comparable to the above-canal settlements. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A14: Comparison to distant settlements for non-farm outcomes (coarsened exact matching robustness specifications)

	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Panel A. Coarsened exact matching, no outliers dropped					
Below canal	0.113***	0.003	0.004	-0.001	0.006
	(0.028)	(0.004)	(0.002)	(0.002)	(0.013)
Above canal	0.039	0.007*	0.004**	0.001	-0.007
	(0.025)	(0.004)	(0.002)	(0.002)	(0.009)
Control group mean	5.534	0.089	0.054	0.024	9.656
Observations	44,242	41,150	41,150	41,150	42,076
R^2	0.30	0.10	0.08	0.13	0.39
Panel B. Coarsened exact matching, 1% outliers dropped					
Below canal	0.113***	0.000	0.003	-0.002	0.006
	(0.040)	(0.005)	(0.002)	(0.002)	(0.011)
Above canal	0.074**	0.006	0.004**	0.000	-0.001
	(0.034)	(0.005)	(0.002)	(0.002)	(0.010)
Control group mean	5.514	0.074	0.049	0.019	9.702
Observations	25,034	23,375	23,375	23,375	23,835
R^2	0.31	0.11	0.08	0.18	0.38
Panel C. Coarsened exact matching, 2.5% outliers dropped					
Below canal	0.088**	0.006	0.003	0.001	-0.003
	(0.035)	(0.006)	(0.003)	(0.003)	(0.012)
Above canal	0.053	0.008	0.004	0.003	-0.004
	(0.033)	(0.006)	(0.002)	(0.003)	(0.012)
Control group mean	5.595	0.082	0.052	0.022	9.691
Observations	16,569	15,372	15,372	15,372	15,729
R^2	0.31	0.10	0.08	0.12	0.37
Panel D. Coarsened exact matching, 5% outliers dropped					
Below canal	0.085*	-0.005	0.003	-0.004	-0.016
	(0.047)	(0.006)	(0.003)	(0.004)	(0.014)
Above canal	0.050	0.005	0.007***	0.000	-0.020
	(0.043)	(0.006)	(0.003)	(0.003)	(0.012)
Control group mean	5.489	0.073	0.050	0.016	9.711
Observations	9,312	8,666	8,666	8,666	8,907
R^2	0.32	0.10	0.08	0.11	0.37

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for non-farm outcomes using a range of samples to test the robustness of our main results presented in Table 5. Here we employ coarsened exact matching (CEM) following Iacus et al. (2012), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions, instead of entropy balancing to ensure the distant settlements are comparable to the above-canal settlements. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A15: Comparison to distant settlements for outcomes disaggregated by land ownership (coarsened exact matching robustness specifications)

	Consumption (log, landless HHs)	Consumption (log) (log, land-owning HHs)	Middle school ed. (share of landless pop.)	Middle school ed. (share of land-owning pop.)
Panel A. Coarsened exact matching, no outliers dropped				
Below canal	-0.010	0.019	0.018***	0.035***
	(0.011)	(0.014)	(0.006)	(0.008)
Above canal	-0.018*	-0.002	0.008	0.014**
	(0.009)	(0.011)	(0.005)	(0.006)
Control group mean	9.516	9.762	0.256	0.348
Observations	40,600	40,869	40,522	40,906
R^2	0.31	0.40	0.35	0.47
Panel B. Coarsened exact matching, 1% outliers dropped				
Below canal	-0.017	0.023**	0.011*	0.030***
	(0.014)	(0.012)	(0.006)	(0.008)
Above canal	-0.022*	0.007	0.006	0.011
	(0.012)	(0.011)	(0.005)	(0.007)
Control group mean	9.543	9.807	0.259	0.353
Observations	23,024	23,189	22,980	23,215
R^2	0.30	0.39	0.39	0.49
Panel C. Coarsened exact matching, 2.5% outliers dropped				
Below canal	-0.023	0.012	0.018**	0.021**
	(0.017)	(0.013)	(0.008)	(0.010)
Above canal	-0.028	0.002	0.012*	0.004
	(0.018)	(0.012)	(0.007)	(0.009)
Control group mean	9.534	9.778	0.257	0.358
Observations	15,188	15,339	15,157	15,353
R^2	0.30	0.38	0.37	0.47
Panel D. Coarsened exact matching, 5% outliers dropped				
Below canal	-0.013	0.021	0.008	0.024**
	(0.017)	(0.016)	(0.010)	(0.010)
Above canal	-0.021	0.012	0.000	0.011
	(0.015)	(0.015)	(0.011)	(0.009)
Control group mean	9.561	9.798	0.267	0.352
Observations	8,589	8,667	8,569	8,671
R^2	0.33	0.36	0.37	0.45

$^{*}p<0.10, ^{**}p<0.05, ^{***}p<0.01$

Notes: This table reports the spillover analysis estimates following Equation 5.2 for various outcomes disaggregated by land ownership using a range of samples to test the robustness of our main results presented in Table 5. Here we employ coarsened exact matching (CEM) following Iacus et al. (2012), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions, instead of entropy balancing to ensure the distant settlements are comparable to the above-canal settlements. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers. The below-canal (directly treated) and above-canal (indirectly treated) settlements are compared to distant settlements far from the canal within the same district. Distant settlements are defined as settlements more than 15km away from a canal. The γ_1 (above-canal) and γ_2 (below-canal) estimates are reported here, along with their significance and standard errors below in parentheses. The control group mean, referring to the area-weighted mean of the above-canal settlements, is reported along with the number of observations with non-missing data for each outcome variable and the adjusted R^2 of the estimate.

Table A16: Effect of canals on town growth (robustness specifications)

	Town existence (pop. 5,000)			Population (log)			Growth (decadal)		
Panel A. 10km radius									
Command area in town catchment area	0.036***			0.068***			0.075***		
(binary treatment)	(0.014)			(0.026)			(0.020)		
Share of 0-10km band in command area		0.087***	0.129**		0.229***	0.183		0.058***	0.054
(continuous treatment)		(0.023)	(0.064)		(0.048)	(0.134)		(0.020)	(0.049)
Share of 10-20km band in command area			-0.052			0.057			0.004
			(0.065)			(0.147)			(0.051)
Observations	21,444	33,780	33,780	21,444	33,780	33,780	19,657	30,965	30,965
R^2		0.68	0.68		0.83	0.83		0.06	0.06
Panel B. 20km radius									
Command area in town catchment area	0.040***			0.085***			0.043**		
(binary treatment)	(0.014)			(0.026)			(0.020)		
Share of 0-20km band in command area		0.090***	0.119*		0.261***	0.236*		0.072***	0.175***
(continuous treatment)		(0.022)	(0.063)		(0.051)	(0.135)		(0.024)	(0.060)
Share of 20-40km band in command area			-0.039			0.034			-0.139**
			(0.070)			(0.155)			(0.063)
Observations	21,636	46,932	46,932	21,636	46,932	46,932	19,833	43,021	43,021
R^2		0.67	0.67		0.82	0.82		0.06	0.06
Panel C. 30km radius									
Command area in town catchment area	0.040***			0.086***			0.050**		
(binary treatment)	(0.014)			(0.027)			(0.020)		
Share of 0-30km band in command area		0.082***	0.065		0.284***	0.242**		0.053**	0.202***
(continuous treatment)		(0.020)	(0.051)		(0.052)	(0.112)		(0.024)	(0.056)
Share of 30-60km band in command area			0.024			0.059			-0.210***
			(0.060)			(0.129)			(0.065)
Observations	21,372	59,016	59,016	21,372	59,016	59,016	19,591	54,098	54,098
R^2		0.67	0.67		0.82	0.82		0.06	0.06
Panel D. 40km radius									
Command area in town catchment area	0.024			0.055*			0.012		
(binary treatment)	(0.017)			(0.029)			(0.020)		
Share of 0-40km band in command area		0.096***	0.019		0.318***	0.274***		0.024	0.123**
(continuous treatment)		(0.021)	(0.044)		(0.053)	(0.094)		(0.025)	(0.051)
Share of 40-80km band in command area			0.109**			0.062			-0.143**
			(0.054)			(0.107)			(0.061)
Observations	20,016	67,488	67,488	20,016	67,488	67,488	18,348	61,864	61,864
R^2		0.67	0.67		0.81	0.81		0.06	0.06

$^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Notes: This table shows the effect of canal construction on town growth reported in Table 6 using various catchment area radii around the town results. An additional analysis includes the command area coverage of a second, outer band around each town as an additional independent variable. Panel B, using the 20km radius to define the town catchment area, is presented in Table 6.

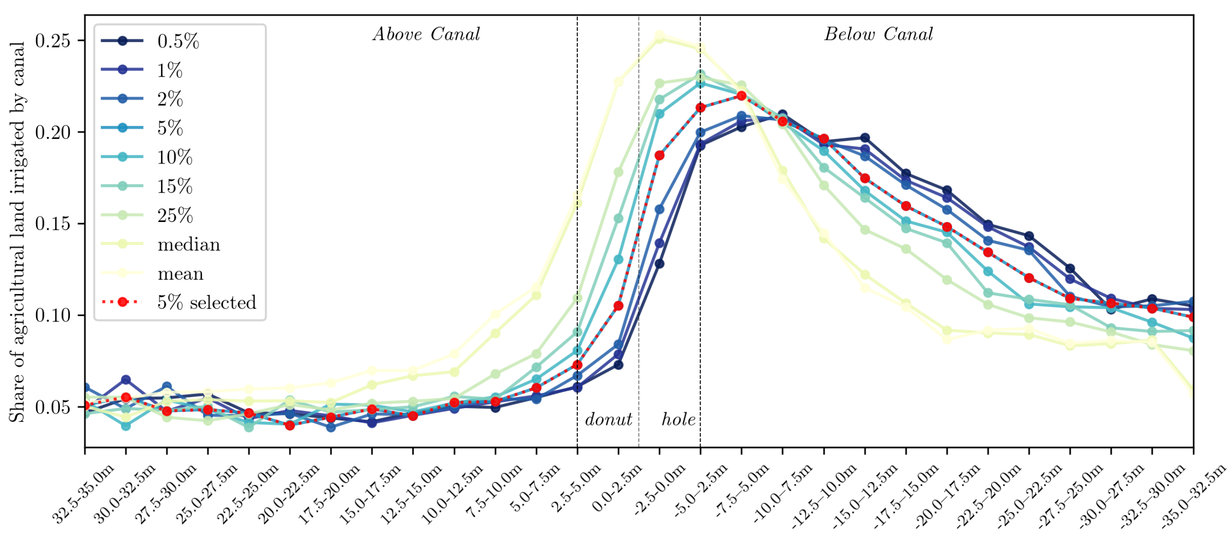
Table A17: Town analysis split by size

	Population (log)			
	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Population:	*< 7,428*	*7,429-13,742*	*13,743-27,365*	*27,366-1,116,3116*
Command area in town catchment area	0.133	0.323**	0.271**	0.089
(binary treatment)	(0.090)	(0.129)	(0.137)	(0.062)
Observations	357	356	357	356

*$p<0.10$, **$p<0.05$, ***$p<0.01$

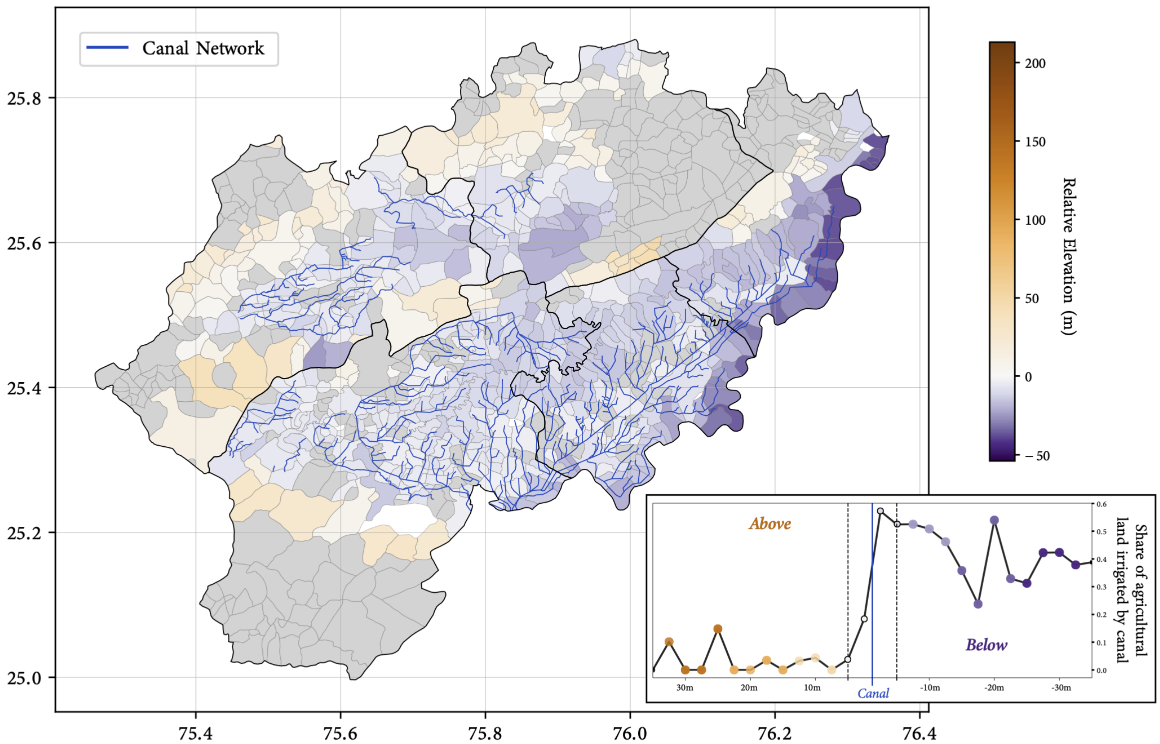
Notes: This table shows the estimated effect of canal area on town growth by town size. Results follow Equation 5.3 and define the binary treatment in the same was as results in Table 6, where a town is considered to be treated when 20% of its catchment area (i.e., within a radius of 20km) has been covered by a command area. Towns are grouped into quartiles using the town population at the time of treatment, with the population ranges used to define each quartile shown in the table. A regression using the log of town population as the outcome variable is run for each quartile independently, with the β_1 for each regression reported here.

Figure A1: Calculating the relative elevation of each settlement



Notes: Each line in this figure uses a different moment of the distribution of elevation in a settlement polygon to define the relative elevation between that settlement and the nearest canal. The elevation of the nearest canals is parameterized by the elevation of the single closest point. Share of agricultural land irrigated by canal is on the y-axis. Relative elevation is plotted on the x-axis, with negative relative elevation indicating settlements below the canal. We select the 5th percentile to define settlement elevation.

Figure A2: Relative elevation RDD empirical strategy



Notes: This figure illustrates our relative elevation empirical strategy using Bundi district in Rajasthan. Each polygon is a settlement (village or town), with its elevation relative to the nearest point on the nearest canal colored orange for settlements above the canal and purple for those below. Settlements that are more than 10km away from the nearest canal (in distance) or within ±2.5m (in elevation) of the nearest canal are excluded (light gray on the map). The inset plots the share of agricultural area that is irrigated by canal vs. the relative elevation for each settlement. The discontinuity is clear, with settlements topographically above the nearest canal having a significantly larger share of canal-irrigated area.