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

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

March 2, 2022

Please click here for the latest version.

Abstract

How do improvements in agricultural productivity translate into development and structural transformation? An extensive literature has addressed these questions, dating back to the earliest days of development economics. In this paper, we estimate the long-run impacts of India's irrigation canals—which span over 300,000 km and deliver water to over 100,000 villages—on structural change and long-run development. Drawing on high-resolution data on every household, firm, village, and town in India, we use three identification strategies to characterize the long-run direct and spillover effects of large increases in agricultural productivity. First, we exploit the gravity-driven nature of canal irrigation in a regression discontinuity design with elevation as the running variable. Second, we study spillovers by comparing untreated (i.e. above-canal) settlements close to canals to those farther away. Third, we use a 100-year panel of urban populations to estimate the effects of canals on regional urbanization. In the long run, canal access drives substantially higher irrigation intensity and land productivity. These changes result in higher population density, but treated areas experience no structural transformation: there are no changes in the share of the workforce outside of agriculture, or even in agroprocessing. Consumption gains accrue only to landowners; we estimate a tight null effect on the consumption of the 60% of the population with little or no land. Structural transformation does occur, but through the growth of nearby towns rather than through sectoral reallocation of labor within treated villages or their near neighbors. Our findings are consistent with a model where labor is mobile in the long run, and where non-farm activity tends to cluster in towns rather than to be spatially dispersed across rural landscapes. 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.

*We are grateful for persistent, patient, and creative 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, the University of Chicago, IIM Ahmedabad, IFPRI, BREAD, Dartmouth College, the World Bank, STEG, and NEUDC. 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.

†Johns Hopkins SAIS, sasher@jhu.edu

‡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 (Lewis, 1954; Johnston and Mellor, 1961; Mellor, 1986; Schultz, 1964). Authors such as Johnston and Mellor (1961), echoed later by Mellor (1986) and Timmer (1988), 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 invoked 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] A critique of these models is that they depend heavily on a closed economy assumption. Matsuyama (1992) showed that in an open economy, increases in agricultural productivity can cause *specialization* in agriculture via comparative advantage.[2]

When testing these theories empirically, the extent to which goods and labor are mobile will depend on the geographic scale of analysis and the time horizon studied. This paper is concerned with understanding the long run effects of technical change in agriculture on the broader economy, with a particular focus on how individuals move across sectors and across space at different geographic scales.

We study the impacts of 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—essentially artificial rivers that carry water into dryland areas—spans over 300,000 km and serves over 130,000 villages, nearly one in four in India. Canals have been especially important in enabling irrigation during the relatively dry winter cropping season. Canals were historically the most important source of irrigation in India, and even in the 21st century they are the second largest

[1] Schultz (1953) referred to this phenomenon as the "food problem". The same mechanism lies at the heart of more recent work, which relied on non-homothetic preferences as the main driver of structural transformation (Gollin et al., 2002, 2007; Alvarez-Cuadrado and Poschke, 2011). 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] Bustos et al. (2016) also use an open economy framework, where agricultural productivity growth leads to an expansion of agricultural output, but whether this increases or reduces the use of labor in agriculture depends on the factor bias of technical change.

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.[3]

These canals are an ideal context for studying the long-run impacts of technical change in agriculture for two reasons. First, they drive large and sustained differences in productivity across otherwise similar locations. Second, they were built many decades ago: half before 1981 and many before 1900. Other agricultural interventions tend to diffuse across space over time, making it more difficult to study their long run impacts.

Studying the effects of canals on the geography of structural transformation requires detailed, high resolution data. We combine microdata from economic and poverty censuses, administrative records, geospatial datasets, and satellite imagery to measure irrigation, agricultural practices and productivity, 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. therefore study the long-run impacts of India's canal network. We are unable to estimate short-run effects as most canals with more recent completion dates appear to have been built many decades ago but have been rehabilitated more recently.

We can think of canals as having effects at four different geographies: (i) direct effects in the locations that they serve with surface irrigation; (ii) indirect effects in nearby unserved locations; (iii) effects in regional urban markets; and (iv) effects at much broader geographies that could diffuse across the entire country or world. We employ 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.

To measure direct effects on canal-irrigated areas, we use a regression discontinuity (RD) design that exploits the gravity-driven nature of canal water distribution, with elevation relative to the nearest canal as the running variable. Canal placement is determined by engineering specifications

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

and topography, and water from canals only flows downhill, treating locations topographically below it. Locations a short distance away from a canal but only a few meters higher than the canal will thus experience no irrigation benefit and can serve as a control group for the irrigation treatment.

The RD analysis tests for long-run differences between below-canal settlements and above-canal settlements. This comparison does not capture the full effects of canals if there are spillovers into above-canal settlements. In the presence of local labor or goods market linkages, untreated settlements above the canal could increase demand for both agricultural and non-agricultural labor. Alternately, canals could recharge regional groundwater tables, increasing access to pump irrigation above the canal.

To measure these more diffuse effects, we compare settlements directly above canals to settlements that are in the same district but are more distant from canals, using entropy balancing (Hainmueller, 2012) that reweights distant settlements to ensure a comparison group with similar distributions of natural characteristics (climate, topography, and agricultural potential).[4]

We expect most spillovers to decay with distance; estimating differences between untreated settlements near and far from canals thus provides a test of the existence of spillovers, as long as they are smaller for more distant villages.[5] Finally, to capture concentrated effects of canals in nearby urban areas, we use a hundred year panel of town populations, the only high resolution panel data available in the era of canal construction. 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 RD 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.[6] The yield effects are observed entirely in the relatively dry winter (*rabi*) season: canals improve water access in a second cropping season but generate no significant differences during the summer (*kharif*) growing

[4]We show robustness to using an alternate methodology for reweighting distant settlements, coarsened exact matching (Iacus et al., 2012), which discretizes continuous geographic variables and reweights according to these coarsened bins, discarding distant settlements that do not fall into bins containing canal settlements.

[5]If spillovers extend frictionlessly to the entire country, then they are impossible to measure given the kind of cross-sectional data that is available.

[6]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.

season, when monsoon rains generally provide sufficient water. There are no spillover effects on agricultural outcomes: irrigation levels, yields, and land use in above-canal settlements are highly similar to those in more distant settlements.[7] The sharp differences in agricultural outcomes between above-canal and below-cabal settlements have been sustained over many decades, making them a useful natural experiment for studying what happens to the rest of the economy after large-scale gains to agricultural productivity.[8]

Turning to measures beyond agriculture, we find that the agricultural changes brought about by canals cause substantial population growth but ultimately little local structural change. Below-canal settlements have sharply higher population density, with no measurable 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 agroprocessing. There is evidently an increased demand for labor (since population density is higher in canal-irrigated areas), but these highly agricultural settlements are not developing substantial non-farm sectors. Our town panel analysis, however, finds concentrated population gains in proximate urban areas in the decades following nearby canal construction. We do not observe the historical sectoral allocation of work in the town panel, but can observe that towns are much more non-agricultural than villages, so that town growth implies at least a local shift in the economy toward non-farm work. A back-of-the-envelope calculation suggests that the net population gain is considerably higher in rural than urban areas, and thus that canals increase the agricultural intensity of the regional workforce, despite inducing some urban growth.

Canals have heterogenous effects on living standards. We estimate consumption for every household in India using small area estimation Elbers et al. (2003). We find canals produce no significant gains to estimated consumption for the 60% of households who own little to no land. In contrast, households with landholdings above 1.32 acres are wealthier, with effects increasing in household

[7]Notably, above-canal settlements do not have higher irrigation from any source, including groundwater. This rules out the possibility of spillovers through a groundwater recharge channel.

[8]The RD results are robust to a range of alternate specifications, including an alternate regression discontinuity using distance to the officially designated command area boundary of the canal. The command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal. This definition exploits finer details of local topography, but risks endogeneity if command area boundaries were drawn such that they intentionally include or exclude certain locations.

land. We do find evidence of some consumption spillovers into above-canal settlements, but only for landed households and with much smaller effects than for landowners in below-canal settlements.[9]

We interpret our results in the context of a multi-sector, multi-location model that is closely related to the prior literature (Matsuyama, 1992; Bustos et al., 2016), but that captures two key features of our context. First, we model labor as fixed in the short run and 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 our 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. Returns to land, the fixed factor, remain higher even in the long run. Structural transformation occurs 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 industrialization.[10] Foster and Rosenzweig (1996) and Foster and Rosenzweig (2004b) studied the impact of the Green Revolution on structural change and educational attainment in a panel of villages across India, finding that agricultural gains raised wages and inhibited industrialization. In the US, Hornbeck and Keskin (2015) estimated only short run benefits to increased access to irrigation due to the tapping of the Ogallala Acquifer, with no long-run structural transformation at the country level. Bustos et al. (2016) found that the introduction of a second harvesting season for maize in Brazil depressed structural change in Brazilian municipalities, whereas the spread of genetically modified soybeans, which represented labor-augmenting rather than land-augmenting technical change, increased the exit of labor form agriculture. Similar to the effects we find on urban growth, Bustos et al. (2020) found that agricultural productivity gains drove urbanization through the flow of land rents to

[9]These spillovers could be due to improved access to irrigation that we cannot detect in our irrigation measures, or increased demand for the types of labor or goods produced disproportionately by landed households. For example, landowners in above-canal settlements may be more likely to own firms that serve the increased demand from below-canal settlements. The limitations of our data prevent us from delving further into this result.

[10]There is also a large body of evidence on responses to transient agricultural productivity shocks due to weather. Emerick (2018) finds that non-tradeable employment increases in districts experiencing positive agricultural productivity shocks, consistent with our model of demand-driven structural change. Adhvaryu et al. (2013) and Colmer (2021) show that flexible labor markets are key to helping workers adapt to negative temperature shocks. This paper speaks less to this literature on transient shocks because we study how people adjust to large, permanent changes in agricultural productivity.

cities.[11] Our study of the direct and indirect effects of canals at different geographic scales may help to resolve some of the differences in this literature. Our empirical and theoretical results show that we should expect differential effects depending on whether areas are rural, where the direct comparative advantage effects of agricultural productivity gains may outweigh demand or capital channels that would induce nonfarm growth, while in proximate urban areas the opposite is likely to be true. We study much smaller geographic units than those in much of the literature; it is possible that county- and municipality-level effects in the US and Brazil are the result of both the direct and indirect effects that we document at high resolution.[12]

We also contribute to the literature on how labor flows respond to economic shocks. A rich body of research documents how both temporary (Imbert and Papp, 2020) and permanent (Greenstone et al., 2010; Allcott and Keniston, 2018) migration respond to economic shocks in both high- and low-income countries. It is notable that much of the prior 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 much 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 much longer-run analysis suggests that it 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.[13]

Our results also generate further evidence on the high barriers to rural industrialization. Asher and Novosad (2020) and Burlig and Preonas (2021) find that major investments in rural roads and electrification respectively have limited effects on non-farm activity in India.[14] Faber (2014)

[11] An example of this capital channel was discussed at length in the context of colonial Bengal in Bose et al. (1993).

[12] Our results also echo the predictions of Foster and Rosenzweig (2004a), which suggested that agricultural productivity shocks have substantially different effects on landowners and the landless, consistent with our findings.

[13] While there is a widespread idea in the literature that permanent migration in India is rare, this claim appears to arise from 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. However, we show that there are at least no contemporary effects of canals on fertility or mortality, suggesting some combination of higher in-migration and reduced out-migration.

[14] Asher and Novosad (2020) find that the main impact of roads is to provide access to non-agricultural labor markets

actually finds that highway construction through peripheral areas in China caused deindustrialization. Infrastructure investments in rural areas may improve well-being and may motivate in- and out-migration, but are unlikely to cause substantial changes in *in situ* nonfarm 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 cities, rather than in the rural areas directly experiencing agricultural productivity gains.

Finally, our work adds to a growing literature estimating the impacts of changes to access to irrigation. Sekhri (2014) shows that access to groundwater irrigation lowers poverty. Blakeslee et al. (2021b) find that the drying up of wells actually 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, Blakeslee et al. (2021a) study canals in India using a command area distance RDD.[15] 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.[16]

Our paper proceeds as follows. Section 2 provides background on India's irrigation canals. Section 3 describes our conceptual framework for understanding how agricultural productivity gains spill over into other economic outcomes. Section 4 describes the data sources and Section 5 lays out our multiple empirical strategies. Section 6 presents our results and Section 7 discusses their interpretation. Section 8 concludes.

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

outside the village. This result is suggested by our model, where towns have productivity advantages for non-farm work.

[15] This is analogous to the secondary identification strategy we use to show robustness of the primary elevation-based RDD.

[16] Blakeslee et al. (2021a) differ from us in arguing that canals *reduce* city size directly in their command area. Our analysis uses a time series of urban populations and finds that new towns emerge in the vicinity of canals and that canals boost overall population in treated areas.

flow irrigation through canals of various types. It was estimated at that the end of the 19th century, India's 12 million hectares of irrigated land amounted to 4 times that of the United States and 6 times that of 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 yield variety 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. By 2011 it had spent more than $7.5 billion to help finish stalled large-scale irrigation projects (Shah, 2011). More recently, states such as Madhya Pradesh have improved canal irrigation by increasing maintenance, distributing water more equitably, investing in last mile distribution networks, reducing political interference, and building cooperation with farmer organizations (Shah, 2011). According to the most recent estimates, canals still account for approximately 1/4 of the net irrigated area in India (Jain et al., 2019), although estimates vary according to the methodology.

Figure 1 shows the distribution of completion dates of India's major and medium canals, which are the focus of this study.[17] 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 2012, the main year in which we measure outcomes, 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 1968 and completion year of 1981. Given that our primary outcomes are measured in 2011-2013, we thus study the impacts of canals that are typically at least thirty years old.[18]

[17]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.

[18]In fact, median canal age is likely much greater, as our manual validation reveals that official canal construction dates often apply to rehabilitation projects rather than original construction, especially for those reporting completion

3 Conceptual Framework

Our conceptual framework builds on a substantial literature modeling the effects of agricultural productivity change on the non-farm sector (Johnston and Mellor, 1961; Matsuyama, 1992; Bustos et al., 2016; Foster and Rosenzweig, 1996, 2007). Our model is very similar to those of the prior literature, but has two key features that suit it to our empirical question. First, we model how the effects of agricultural productivity change differ in the short and long run; we assume that labor mobility across space is low in the short run and high in the long run. Second, we allow a distinction between villages and towns, where towns are characterized by some kind of agglomeration or trade connections that give them an advantage in producing manufactured goods.

A stylized view is that India's rural economy consists of a large number of predominantly local sub-economies that are embedded in a larger national economy. Each of the local economies features an expanse of agricultural land, divided into villages, typically with a larger market town that serves as an economic center. The villages that surround each market town are mostly small, with an average size of approximately 1000 inhabitants, most of whom work in agriculture. 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 workers from a large pool of landless workers.

These observed features of the data give shape to our conceptual framework. Our model economy consists of two spatial units: villages and towns. A region consists of a single town and its surrounding villages, with their agricultural land. These regions are embedded in a larger economy. Regions are price takers with respect to the broader economy, although we allow for transport costs that imply variation in prices across locations. The economy has two goods: an agricultural good and a non-agricultural composite. Both are tradable, subject to frictions that are described below.

Consumer preferences are defined over the agricultural and non-agricultural goods. As is standard, we assume a non-homotheticity that implies poor consumers will spend larger shares of their income on the agricultural good than richer consumers. Preferences over the non-agricultural good distinguish between goods produced in different locations. Consumers view non-agricultural goods

after 1990.

produced in their own village as closely substitutable with the goods produced in the nearby town. Non-agricultural goods from the rest of the economy, however, are only weakly substitutable with local goods. This assumption reflects the fact that the non-agricultural goods include some that are less than fully tradable (e.g., retail services).

Each village is endowed with an initial population that is divided between landowners and landless workers. For simplicity, we assume that all villages have the same land area. This means that average land holdings (i.e., land per landowner) may vary across villages. Similarly, the initial ratio of laborers to land may vary. Within a region, we assume that land is *ex ante* homogeneous in quality across villages. However, we allow for the possibility that land productivity may change over time, due to the arrival of irrigation canals. Irrigation canals raise agricultural productivity in "treatment" villages but have no direct effect on productivity in other villages or towns.

The agricultural good is produced with land and labor, according to a constant returns to scale technology. As noted, agricultural productivity at the village level depends on the presence of irrigation canals. The non-agricultural good is produced using only labor, according to a technology that displays diminishing marginal returns to labor inputs. Non-agricultural productivity varies across villages and towns. We treat non-agricultural productivity as benefiting from an agglomeration externality, such that productivity depends on the absolute number of people engaged in non-agricultural production in the preceding period. This agglomeration externality implies that towns – which by construction have larger non-agricultural populations than villages – will have an absolute advantage in non-agricultural production, relative to their surrounding villages. However, the diminishing marginal returns to labor imply that villages will undertake some of their own non-agricultural production activity.

We are interested in the impacts of irrigation canals both in the short run and the long run. As other researchers have argued, labor mobility in rural India faces significant frictions in the short run (Foster and Rosenzweig, 2007; Munshi and Rosenzweig, 2016) although Foster and Rosenzweig (2007) find male out-migration from villages of over 20% when considering a longer (17 year) period. Our conceptual framework thus considers three distinct time periods: a baseline before the construction of the canals; a short run after canals have been constructed but before labor has adjusted to the changes in

productivity; and a long run in which a spatial equilibrium holds, and in which real wages are equalized across locations. This long-run labor mobility is a key feature of our theoretical framework and also of our empirical work. Allowing for labor mobility leads to important differences relative to prior models.

In our framework, positive agricultural productivity shocks, such as the investments in irrigation canals studied in this paper, will lead to short-term growth in local demand for land and labor, driving up wages and land rents, and leading to higher incomes in treated communities for both landowners and landless workers. However, in the long run, the higher returns to labor are dissipated due to an influx of workers. Real wages equilibrate for labor, as the mobile factor, and the treated communities end up with higher population density. Returns to land, the fixed factor, remain higher even in the the long run.

The higher incomes of landowners, along with the increased population, result in higher demand for non-agricultural goods. This is amplified to some degree by the standard non-homotheticity of preferences, which implies that the landowners spend their increased rents disproportionately on non-agricultural goods. (In the transition to the spatial equilibrium, there is also a short-run increase in wages before population growth brings them into spatial equilibrium with the outside world.)

As in standard models, the increased demand for non-agricultural goods drives structural transformation (an increase in the share of workers in the non-agricultural sector). But since our local economies are open to the rest of India, the structural change need not take place within the same communities that benefit from gains to agricultural productivity. In a frictionless world, the increased demand for non-agricultural goods could be met by production anywhere. Our model thus takes no stance on the locations that will experience structural change. We allow for increased production of non-agricultural goods to take place within villages, nearby towns, and/or the rest of the Indian economy.

What determines the location of structural change? In other words, what determines where we see a take-off in non-agricultural production activity? In our spatial framework, each local economy consists of a set of villages that are connected, in a hub-and-spoke arrangement, to a central town. Trade between villages and towns faces a transport cost friction. Towns are connected to one another (and to the external Indian economy) frictionlessly. This setup captures the reality that movement

between villages is costly, and towns function as central meeting and marketing places. Within the local economy, this means that a sale of goods from one village to another would accrue a double transport cost, relative to a sale between village and town. The model implies that increasing demand for non-agricultural goods within a region will be met by some combination of production in villages and in their central town, with the precise allocation depending on the relative productivity levels, the substitutability of town goods and village goods, and the transport costs between towns and villages. When town productivity is high and transport costs are low, and town goods are highly substitutable with village goods, we expect to see increases in demand met through production in towns. At the other extreme, when town productivity is only slightly higher than village productivity, and transport costs are high, increases in demand might be met through production in villages. The model permits both possibilities; it is an empirical question for this paper to see how increases in demand for non-agricultural goods may lead to structural change in different locations.

4 Data

To estimate the impacts of canal irrigation on local economic outcomes, we assemble recent high-resolution data on the universe of firms, households, and settlements in India, building on data from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG) (Asher et al., 2021). Because the reclassification of (rural) villages into (urban) towns is an endogenous outcome based on population density and administrative discretion, we use a pooled dataset where the unit of observation is a town or village (as provided by the SHRUG), which we call a settlement henceforth.[19]

The 2011 Population Census provides demographic variables along with village-level data on cultivated and irrigated land area in every village in India. It also records the main three crops grown in each village, from which we create an indicator variable for villages that grow a water-intensive crop (cotton, sugarcane, or rice). Since settlements vary in their physical area, our preferred measure for population is population density, which we measure as inhabitants per km^2.[20]

The 2012 Socioeconomic and Caste Census (SECC) is an asset census undertaken in all of India to

[19]Note that data on agricultural outcomes are generally available only in villages.

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

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 a small area estimate of predicted consumption per capita, which is a linear combination of data recorded in the SECC. The latter follows the estimation methodology of Elbers et al. (2003) and is described in detail in Asher et al. (2021).[21] 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 nonfarm economic establishments in India, which we use to measure non-agricultural economic activity for each settlement. We calculate employment as a share of the adult population, as recorded in the 2011 Population Census.[22] We use the National Industrial Classification codes of firms in the Economic Census to calculate the share of the adult population specifically employed in manufacturing, services, and agroprocessing.[23]

In the absence of directly-measured settlement-level agricultural productivity data, we use the Enhanced Vegetation Index (EVI), a satellite-derived measure of biomass with 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 the productivity value by subtracting the mean of the first six weeks of the season from the maximum 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.[24]

[21] For a secondary measure of educational attainment, we use the settlement literacy rate from the Population Census.

[22] 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.

[23] 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. Agroprocessing is defined as a subset of manufacturing employment codes, specifically NIC codes 10 and 12.

[24] We find similar effects if we use different years, which is expected given that we are studying equilibrium effects of canals, and similar effect significance if we use EVI levels rather than logs. See Asher and Novosad (2020) for

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.[25] A canal command area is the area determined by canal engineers for which canal irrigation is feasible. The command area begins sharply at the canal (as areas above the canal cannot be gravity-fed) and ends along a threshold 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 of individual canals suggested that recent start and end dates in WRIS often represented canal rehabilitation rather than new canal construction.[26] It is therefore challenging to identify exact construction dates of recent canals. Older construction dates appear to be more credible, as canal investments in the independence period and earlier were more often targeted at new canals than at maintenance.

Using settlement polygon GIS data from MLInfoMap, 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, town, and coast.

Table 1 reports means for all variables used in the analysis. There are 590,855 settlements (villages and towns) in our dataset that contain non-missing population data. 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.

more details on construction of this measure.

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

[26]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. Note that 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.

5 Empirical Strategy

Testing for the long-run impacts of increasing agricultural productivity is challenging for two primary reasons. First, the placement of infrastructure such as canals is likely to be endogenous: large, costly infrastructural investments such as canals were plausibly targeted to areas with political favor and high returns to irrigation. We use three empirical strategies to overcome these challenges. 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 onto nearby untreated locations, we comparing 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 areas, 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 trenches, pipes, and secondary canals. 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 be able to access canal water. Our main identification strategy thus 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 can be thought of as controls.

A settlement polygon is characterized by a set of pixels with a distribution of elevation values. We define the polygon elevation as the 5th percentile of the polygon pixel distribution; this value strongly predicts the difference in canal irrigation between treatment and control areas (see Appendix Figure A1).[27] To calculate canal elevation, we select the elevation of the nearest point on a canal to a given settlement.

Equation 5.1 describes the regression discontinuity design (RDD) specification, following Imbens

[27] Results are similar if we use the 25th or median elevation to define above/below canal thresholds (see Appendix Tables A3, A4, A5, and A6). We chose the 5th percentile in order to have a control group with close to zero canal irrigation; when we estimate spillover effects below, it is particularly desirable for the above-canal group to experience no direct treatment by canal water.

and Lemieux (2008) and Gelman and Imbens (2019):

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

where $y_{i,s}$ is an outcome in settlement i and subdistrict s and $REL\_ELEV_{i,s}$ is canal elevation minus settlement elevation (such that a positive 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, distance to the nearest river, distance to the coast, and the GAEZ crop suitability measure for irrigated rice and wheat.[28] ν_s is a subdistrict fixed effect, which 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.

The analysis sample includes settlements less than 10km of distance and 50m of vertical elevation from the nearest canal.[29] We limit the sample to subdistricts that have at least one settlement in the treatment group and one settlement in the control group. Settlements 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 treated) and some of its land below the canal (and thus treated). 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

[28] We use these as our best measures of overall 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.

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

entire subdistrict is dropped from our sample. Table 1 shows the sample size and summary statistics for each subset of the data. We use the ruggedness-balanced sample 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 nonfarm job for every 10 adults, and just under half of adults have completed primary school.

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 largely built in the 19th and early-mid 20th centuries and treatment status is determined at the settlement level, there are no available high-resolution socioeconomic or agricultural data to test this assumption. However, we can test for differences in time-invariant geographic 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 geographic 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, or crop suitability for rice or wheat. We do estimate that below-canal areas receive 3.6mm *less* annual rainfall on average (on a mean of 1167mm). This effect is tiny in magnitude, would have the opposite effect on agricultural productivity as canal irrigation. We also find that treatment settlements are approximately 10% closer to a river than control settlements, but this should not bias our estimates as distance to river and all other geophysical fundamentals are included as controls in all specifications.

As a robustness check, we use a secondary regression discontinuity design that compares settlements just inside and just outside of the canal command area.[30] We define the running variable as the distance between settlement centroid and command area boundary, defining it negatively outside of the canal.[31] 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

[30]This is similar in design to the strategy used in concurrent work by Blakeslee et al. (2021a).

[31]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.

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 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. 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 as within the command area.[32] We test for balance with this command area boundary strategy in Table A2, finding no evidence for any imbalance.

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 areas. 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 areas are part of integrated labor markets, then labor market effects of canal irrigation could be expected to diffuse across the treatment boundary; if labor mobility was high enough, we could estimate zero differences between these areas in the RDD analysis above even in the presence of true effects. Alternately, enhanced agricultural productivity below canals could motivate structural change not only below the canals but also above, attenuating the RDD estimate. More directly, canals could recharge 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. This strategy is predicated on the assumption that any mechanism driving spillovers is likely to decay with distance from treated areas. These settlements are 5km 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. If spillovers do not decay over this distance, they are more difficult to measure. For example,

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

if landless labor is 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 non-zero spillovers as long as they have a non-zero gradient in distance, but they will be underestimated to the extent that those spillovers extend into the distant control group.

Our preferred specification employs entropy balancing (Hainmueller, 2012) to assign weights to distant settlements so that the distributions (first, second, and third moments) of all geophysical variables in distant settlements match the distributions of these variables in above-canal settlements. This method is increasingly favored over other methods because it does not require functional form assumptions on the propensity weights and thus achieves better balance than propensity-score matching methods.[33] Following the literature, we enforce common support by dropping outliers (top and bottom 2.5% for each of the geophysical variables). As a backup strategy, we use coarsened exact matching (Iacus et al., 2012) to define and weight a matched sample of distant villages to above-canal (control) settlements. This method coarsens our geophysical variables into discrete bins and matches settlements on all coarsened variables, discarding distant settlements who do not perfectly match any above-canal settlements on geophsyical characteristics and reweighting the rest to match the weights of the matched 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 correspond exactly to the set of settlements used 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 subdistrict 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

[33] See Athey and Imbens (2017) for a discussion of how this and similar methodologies seek to create comparable treatment and control groups in the absence of exogenous variation. For recent examples of empirical work using entropy balancing, see Basri et al. (2021) and Guriev et al. (2021).

into above-canal settlements, we expect γ_1 to be non-zero.

Note the difference between γ_2 here and the RDD estimate 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.

5.3 Town growth through time

The empirical strategies thus far estimate differences between canal irrigated settlements, proximate unirrigated settlements, and similar settlements farther away. But they do not capture the possibility of structural change through concentrated urbanization in areas with economic linkages to canal zones, for two reasons. First, a town may be affected by increases in regional agricultural productivity even if is not in or near the irrigation zone. Second, the spillovers analysis above estimates average effects and is not well suited to test for concentrated changes in a small number of urban locations. As suggested by our model, agricultural productivity is likely to generate disproportionate growth in locations with productivity advantages in non-agricultural production, such as those with natural advantages or agglomeration economies. These advantages may not require being in or adjacent to the canal-irrigated zone.

To test whether canals affect regional urbanization, we exploit variation in canal construction dates, and examine the emergence and growth of towns in their vicinity. 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.[34] Such an analysis is not possible for any other outcome, because urban population is the only variable available in a long panel spanning many decades of canal construction.

To describe whether a town is near a canal, we first draw a 20 km circle around each town. We define canal treatment for town i in year t as the percentage of the circle area that is overlapped

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

by canal command areas. An alternate specification defines a binary treatment variable that take the value 1 if more than 20% of the 20 km circle is covered by canal command areas. Standard errors are clustered at the district level.

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:

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

where outcome $y_{i,t}$ is either an indicator for town existence or log(town population) and ζ_i and ν_t are town and decade fixed effects, respectively. When $y_{i,t}$ represents population, we define it as the log of 2000 plus the town population — this treats setttlements before they became towns as if their size was just below the average population at which towns first appear in the data.[35] For the binary treatment, we use the estimator from Callaway and Sant'Anna (2020), using the not-yet-treated group as controls and defining treatment to be when a town's 20 km radius catchment area is more than 20% covered by canal command areas.

6 Results

6.1 Direct Treatment Effects of Canals: Regression Discontinuity Estimates

We first report RDD estimates of the direct effects of canal irrigation on agricultural outcomes, the mechanism through which we expect all other equilibrium effects to occur. Panel A of Table 3 shows that in canal-treated areas, 7.4 percentage points more of the land under cultivation is irrigated (17% more than in control settlements), and 9.6 percentage points (300%) more land is irrigated by canals. There are no 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 as suggested by Shah (2011)) and find no effects.[36]

[35] 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 2000 is more likely to represent the population of pre-urban settlements.

[36] Note that the elevation-based "treatment" classification describes something less that full exposure of a settlement to a canal; out of the 133,000 settlements in India with *any* canal irrigation, the median settlement has 45% of its land irrigated by canals.

Panel B in Table 3 reports direct effects of canal irrigation on agricultural outcomes. Canal-exposed settlements experience higher agricultural productivity, with effects concentrated in the relatively dry winter (*rabi*) growing season. Treatment settlements have 7.3% higher values of our satellite-derived land productivity measure in the dry season, and positive but much smaller higher productivity in the wet (*kharif*) season (1.7%, $p=0.058$). This is consistent with the primary role of canals being to improve water access during the dry winter growing season, and being less crucial for productivity in the high precipitation summer. The lack of substantial effect on summer yields is further evidence that our treatment is orthogonal to settlement agricultural potential. Settlements below canals also cultivate 2.9 percentage points more of their total area, a 5% increase over control settlements, and are more likely to grow water-intensive crops. We find no evidence of increased capital intensity of agriculture, as measured by the share of households owning mechanized farm equipment.

The key question of this paper is how these 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, nonfarm employment, and predicted consumption. The only significant effect is on population: by 2011, treatment settlements have 15% 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. Our data do not allow us to observe migration flows, but we reject current positive effects on fertility and mortality (as proxied by share of the population aged 0-6 and 70+), suggesting that this result is driven by net in-migration (Table A1).[37] We estimate a statistically insignificant positive 0.09% effect of canals on average living standards ($p=0.103$): we can reject a 1.8% increase and -0.2% decrease in predicted consumption in below-canal settlements with 95% confidence. We do not find a substantial degree of structural transformation: there is no effect on the nonfarm employment share of the economy, nor do we find significant effects when we disaggregate employment into manufacturing or services. Even when zeroing in on agroprocessing, the sector with teh strongest linkage to agricultural production, we find no effects the employment share (Table A1).

[37] It is possible that in the years following canal construction there were changes to fertility or mortality that we do not measure in our data, but we hypothesize that any previous fertility or mortality effects are likely to persist into the present.

Note that total non-farm employment has risen (as would be expected given the increase in population) but the non-farm share of the economy (the outcome of interest) is unchanged. We do find increases in human capital in canal areas, as shown in Panel D of Table 3. Below-canal settlements show small but precise increases of approximately 1 percentage point in the share of the adult population that has completed primary, middle, and secondary school, as well as the population literacy rate.

Due to the simplicity of the RDD, we can inspect our regression estimates visually. Figure 2 shows regression discontinuity binscatters of key outcomes in each of the categories above, with outcomes residualized on fixed effects and geophysical controls. These figures report the treatment effect at the treatment threshold from Table 3. The binscatters confirm the RDD estimates: clear jumps at zero relative elevation in overall and canal irrigation, *rabi* productivity, population density, and educational outcomes, while no discontinuities are apparent in *kharif* productivity, nonfarm employment, or consumption. 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. Canals lead to a nearly 0.2 SD increase in in total irrigation, driven by a nearly 0.8 SD increase in canal irrigation. The gain in population density is almost 0.2 SD and the increase in educational attainment is nearly 0.1 SD for each of the different measures.

Guided by our theory, we test for the differential returns to land and labor. Our 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 land rents and wages, the returns to these factors can be approximated by estimating canal treatment effects on predicted consumption separately for landless households (who own only labor) and those who own land. We can interpret the consumption effects for landless households as the impact of canals on the returns to labor, and the consumption effects on households with land as the combined impacts on returns to labor and land.

We find that only the returns to land increase relative to above-canal settlements. Panel A of Table 4 shows a decline in the share of the population that are landowners in treatment settlements relative to control settlements, with the average landholding size of landowners unchanged. This implies that

the population increase in canal settlements is disproportionately driven by an increase in the number of landless households. Consumption effects of canals are substantially different for landed and landless households: we find no significant consumption effects for landless households, but landowner consumption is 2.1% higher in below-canal settlements ($p=0.0001$); this result is significantly different from the estimate for landless consumption. Partitioning landowners by land quintile, effects increase (almost) monotonically by quintile, with zero consumption effects on those owning <1.32 hectares of land (the 1st quintile), and a 3.1% effect on consumption for those in the top quintile owning >6.10 hectares (Panel B).[38] Both landless and landowners experience gains in educational attainment, but effects for landowners are two to three times higher than for the landless (Table 4 Panel C).

6.1.1 Robustness

In this section, we show that our results are robust to a range of alternate specifications. First, we show that our results broadly hold up if we omit the sample exclusion restrictions that we put in place to create a more balanced sample. We run our primary estimation on the key outcome variables using the full sample of all canal area settlements, then all canal area settlements with the "donut hole" removed but without imposing any restriction on ruggedness balance. We also estimate these key outcomes with the same restrictions as our main, balanced analysis sample but instead use the median and 25th percentile measures of elevation to parameterize settlement elevation, rather than the 5th percentile used in our main analysis. Next, we estimate the effects of canals 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. While we consider this a secondary identification strategy due to the potentially endogenous drawing of command area boundaries, it is instructive to test whether our results are replicated with a different source of variation.

Appendix Tables A3, A4, A5, and A6 show results for the main outcomes for all of these specifications, with panel F showing our main results for reference. Table A3 confirms that canals cause large gains to canal irrigation but limited to no changes for other forms of irrigation. Table A4

[38] We define quintiles in the landholding distribution based on national data, to maintain consistent quintile boundaries across settlements.

confirms broad agricultural impacts across specifications on both the extensive (share of village land under cultivation) and intensive (land productivity, crop choice) margins. In some samples we also find some evidence of mechanization of agricultural (proxied by share of households owning mechanized farm equipment). Our main non-agricultural estimates remain across all of the different specifications: canals' principal effect is to increase population density and landowner consumption, and there are negligible effects on nonfarm employment shares in canal-irrigated areas. All panels confirm the positive effects of canals on education.

Finally, we test for sensitivity of the main outcomes to parameter choice in the elevation RDD. 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). The numbers in bold in the first column indicate the parameter values used in the primary analysis.

6.2 Estimates of spillovers of canals to above-canal settlements

The results above may not fully describe the general equilibrium effects of canals in the long run. If above- and below-canal settlements are economically integrated, not unlikely given their geographic proximity, then there could be important spillovers. For example, if labor markets are integrated, then landless wages could rise in both above- and below-canal settlements. It is also possible that canals have effects on higher elevation villages through groundwater recharge. The results above are therefore insufficient to reach the conclusion that canals have had no effect on landless wages or in situ structural transformation.

As described in Section 5.2, we study spillover effects by comparing both above- and below-canal regions to more distant regions, matching on and controlling for various natural settlement features. Table 5 shows pairs of coefficients from Equation 5.2 describing the difference between (i) below-canal (treatment) settlements and distant settlements; and (ii) above-canal (control) settlements and distant settlements. In each case, the coefficient shows the estimated effect of being in the canal region relative to distant, untreated settlements.[39]

[39]The difference between the two coefficients is analogous to the RDD treatment effects in Section 6.1, but it represents an OLS estimate of the difference rather than the better-identified RDD estimate presented above.

We focus on the spillover effects ("Above canal" coefficient). We find evidence of small but significant spillovers in agricultural outcomes: above-canal villages have slightly more irrigation and are more likely to grow water-intensive crops relative to distant settlements, although there are no significant differences in total agricultural land or agricultural productivity in either major growing season. We find only limited evidence of non-agricultural effects. No difference between above-canal and distant settlements is significant at the 95% level, although we do find that population density is 3.8% higher in above-canal village (p=0.097), which we interpret as the impact of the limited gains to agricultural outcomes. We provide an interpretation of these results in the following section.

6.3 DiD Estimates of the Effects of Canals on Urban Growth

The evidence presented so far suggests that in this context, there is no broad structural transformation, either in canal-treated settlements or through spillovers to nearby areas, in response to canal-induced agricultural productivity growth. Instead, these agricultural productivity changes are largely absorbed in the long run by net changes in population. However, structural change is typically a geographically concentrated process (Michaels et al., 2012) and closely linked with urbanization, at least in countries not specialized in mineral extraction (Gollin et al., 2016). If regional agricultural productivity improvements caused concentrated changes in settlements with both market linkages to canal areas and high productivity in non-farm activities, our estimates above would likely not have the precision to capture such an effect.

In this section, we use panel data on urban populations in India going back to 1901, in combination with canal construction dates, to test whether town growth responds to nearby canal construction. Figure 5 provides suggestive evidence that towns grow more quickly after canal construction, by showing the log population of the set of town locations that can be observed for 30 years before and after a canal is built within a 20 km radius of the town. These populations are plotted against the decade relative to the date of canal construction, and show a clear increase in the rate of population growth in the 30 years after nearby canals were built than in the 30 years before.

We use a difference-in-differences estimator with a continuous treatment defined as the share of a 20 km radius circle around a town that is in the command area of a canal (Equation 5.3) and

a binary treatment where a town is considered treated when more than 20% of the land in the 20 km radius circle around the town is covered by canal command areas. Results are shown in Table 6. We find that towns are more likely to appear (crossing both the 5,000 and 10,000 population threshold) following nearby canal construction (Columns 1-4). We also find that overall population growth is faster in canal-treated areas, with the estimate from the continuous treatment suggesting that towns entirely surrounded by command areas grow 7.4% faster than similar towns not treated by canals (Column 5), both highly statistically and economic significant. The binary treatment also yields a positive, albeit statistically insignificant, coefficient on town growth (Column 6, p=0.26).

7 Discussion

Our results have two parts. First, differences between areas receiving canal water and nearby areas that don't (the RDD estimates) demonstrate that the direct land productivity effects of canal irrigation are sustained in the long run. Decades after the canals were built, there is a sharp discontinuity in irrigation and agricultural output between canal treatment and control areas. Landless wages (proxied by consumption) are fully equilibrated across this boundary; the sharp difference in population density across the canal boundary suggests that net population movement is the primary factor behind this wage equilization. Landowners in canal areas are better off, presumably benefiting from the higher land rents brought by increased access to irrigation. There is no sign of structural change at this scale; below-canal areas have as few non-agricultural jobs per capita as above-canal areas.

Second, our spillover estimates clarify that these close-range effects of canals are not hiding a broader pattern of structural change in rural areas. One could imagine that canal areas experienced structural change in both above- and below- canal settlements, in which case they might have higher landless wages and different nonfarm industry shares from settlements further away and less plausibly affected by the canal. But we find no evidence of these spillovers: distant settlements have as few non-agricultural jobs per capita as canal treatment or control settlements. Instead, we find evidence for agricultural productivity gains translating into urbanization: towns grow faster following proximate canal construction.

These results are rationalized by our model, where non-agricultural productivity advantages

cause structural transformation to be spatially concentrated. The exact links between growing urban markets and greater agricultural productivity in canal areas are difficult to disentangle in our empirical setting, where historical panel data is only available for town population and not for other outcomes. Bustos et al. (2020) shows that landowners in Brazil invested land rents in urban areas, sometimes at substantial distance, which is a plausible driver of structural change in our setting as well. Clemens (2014) suggests that land rents could be used to finance migration, another channel for urbanization and wealth accumulation among landowners. Due to data limitations, we must stay agnostic as to the exact channel(s) driving the urbanization we document.

The caveat to these spillover result is that it remains possible that linkages propagate impacts of canals to even more distant parts of India, or even the world. For example, if India wer characterized by a single, large, high-mobility landless labor market, then it would be possible that canals raised landless wages equally everywhere, such that distance to the canal has no effect on wages. We expect that this story is at least partially true — if the population change in canal areas was driven at all by migration, then a non-zero labor demand elasticity in the sending locations implies some wage gain there, though it may be too small to detect. We therefore remain agnostic on the aggregate nature of these spillovers; our key result is that the primary equilibrating force following agricultural change was net population movement into treated rural areas to work in agriculture and into regional urban areas to work in the non-agricultural sector. There is evidence that at least some of these spillovers are limited: population density effects fall rapidly away from canal areas, as do consumption gains, and the urbanization effects we find are concentrated a short distance from canals. Consistent with other work on transportation costs in LMICs, distance is likely to remain a substantial friction for many markets (Atkin and Donaldson, 2015).

Some of our results have implications beyond our stylized model of structural change. The increase in educational attainment among both landed and landless fits into a literature examining the relationship between economic change and human capital investments. 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) offer the suggestion that demand for school investment among the wealthier land-rich could have resulted in more schools which ultimately provided benefits to the landless. Our work at present cannot identify the mechanism behind the schooling gains that we observe, so we leave testing between these hypotheses to future work.

8 Conclusion

India's canal systems provide an ideal testing ground for examining the geographic relationship between agricultural productivity improvements and long run structural transformation. Canal irrigation raises agricultural productivity – and especially the returns to land. A unique feature of canals is they create sharp spatial changes in agricultural productivity that can persist for decades after they are built.

In the long run, we find that spatial equilibrium is restored primarily through substantial changes in the size of the laborer population. Decades after the canals were built, there are few differences in living standards between landless workers in canal and non-canal settlements. Nor are there differences in non-farm activity in treated areas. However, structural transformation has taken place, with towns emerging disproportionately near canal-irrigated settlements.

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 locations, changes in fertility, or even 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 during the License Raj period, where manufacturing investments

were slow and state-inhibited and may have had difficulties responding to changes in labor demand, potentially enhancing the role of mobile labor. Whether modern agricultural shocks will be equally mitigated by labor flows remains an important question for researchers.

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 there are tens of millions of temporary migrants on top of those. Our study suggests that this large mobile population is 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*, 3 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.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," 2015. Working paper.

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*, December 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*, 6 2016, *106*, 1320–1365.

_ , Gabriel Garber, and Jacopo Ponticelli, "Capital Accumulation and Structural Transformation," *The Quarterly Journal of Economics*, 5 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 (forthcoming)*, 2021.

Douie, James, "The Punjab Canal Colonies," *Journal of the Royal Society of Arts*, 5 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*, 11 2018, *135*, 488–503.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 7 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*, 9 1996, *86*, 931–953.

_ and _ , "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000," *Economic Development and Cultural Change*, 7 2004, *52*, 509–542.

_ and _ , "Technological Change and the Distribution of Schooling: Evidence from Green-Revolution India," *Journal of Development Economics*, 6 2004, *74*, 87–111.

_ and _ , "Chapter 47: Economic Development and the Decline of Agricultural Employment," *Handbook of Development Economics*, 1 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*, 10 2017.

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*, 6 2021, *129.*

_ , **Remi Jedwab, and Dietrich Vollrath**, "Urbanization with and without industrialization," *Journal of Economic Growth*, 2016, *21* (1), 35–70.

_ , **Stephen L. Parente, and Richard Rogerson**, "The Food Problem and the Evolution of International Income Levels," *Journal of Monetary Economics*, 5 2007, *54*, 1230–1255.

_ , **Stephen Parente, and Richard Rogerson**, "The Role of Agriculture in Development," *American Economic Review*, 5 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*, 1 2020.

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.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding, "Urbanization and structural transformation," *The Quarterly Journal of Economics*, 2012, *127* (2), 535–586.

Mukherji, Aditi, "Evolution of Irrigation Sector," *Economic and Political Weekly*, 12 2016, *51.*

Munshi, Kaivan and Mark Rosenzweig, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 1 2016, *106*, 46–98.

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.

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.

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.

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*, 7 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*, 1 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.

Table 1: Summary statistics

	All India	All canal-area settlements	All canal-area settlements minus donut hole	Balanced analysis sample
Sample Size	589,950	245,531	132,969	91,465
Percent Treatment	–	83	77	80
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)	5.238	5.699	5.514	5.546
Consumption (log)	9.726	9.755	9.750	9.760
Total nonfarm employment (share of adult pop)	0.119	0.105	0.107	0.107
Services employment (share of adult pop)	0.077	0.067	0.067	0.066
Manufacturing employment (share of adult pop)	0.029	0.031	0.032	0.033
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 in our All-India sample that have population data. All canal-area settlements includes towns and villages $\leq$10km from the nearest canal in distance and within 50m of the nearest canal in elevation. Removing the donut hole removes settlements $\pm$2.5m in elevation from the nearest canal. The balanced analysis sample then imposes a balance criteria on ruggedness by dropping all subdistricts in which there is a $\geq$25% difference in average ruggedness between below-canal (treatment) and above-canal (control) settlements. The mean values shown here for the balanced analysis sample also exclude subdistricts that do not contain at least one settlement in each of the treatment and control groups. All averages are weighted by land area.

Table 2: Balance in the RDD using relative elevation

	Ruggedness	Annual rainfall avg. 2010-2014 (mm)	Distance to coast (km)	Distance to river (km)	Wetland rice (GAEZ)	Wheat (GAEZ)
Below canal	-0.008	-3.557**	0.078	-2.390***	0.007	0.000
	(0.056)	(1.694)	(0.371)	(0.425)	(0.011)	(0.004)
Control group mean	4.754	1166.971	362.341	24.562	2.272	0.767
Observations	91,465	91,465	91,465	91,465	91,465	91,465
R^2	0.594	0.988	0.999	0.874	0.923	0.983

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

Notes: This table reports the regression discontinuity estimate for geophysical variables following Equation 5.1, dropping each outcome variable from the list of controls. 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	0.074***	0.096***	-0.005	-0.007
	(0.008)	(0.007)	(0.007)	(0.005)
Control group mean	0.430	0.032	0.210	0.194
Observations	83,182	83,192	83,247	82,385
R^2	0.610	0.380	0.470	0.630

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 households)
Below canal	0.029***	0.017*	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	0.610	0.820	0.700	0.730	0.300

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 per capita (log)
Below canal	0.150***	-0.063	-0.004	-0.056	0.009
	(0.023)	(0.053)	(0.007)	(0.048)	(0.005)
Control group mean	5.282	0.124	0.066	0.049	9.744
Observations	91,465	85,342	85,342	85,342	86,842
R^2	0.450	0.000	0.000	0.000	0.530

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 (share of pop.)
Below canal	0.013***	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	0.570	0.560	0.530	0.580

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

Notes: Results for all outcome variables each separately estimated following Equation 5.1. The β_1 coefficient is reported in the first row for each variable, with the stars indicating its significance and the standard error below in parentheses. The control group mean, number of observations with non-missing data for that outcome, and adjusted R^2 for each regression are each reported.

Table 4: Regression discontinuity results for outcomes disaggregated by landownership

Panel A. Landownership overview

	Landowners (share of households)	Avg. size of land holdings (log hectares, all households)	Avg. size of land holdings (log hectares, land owners)	Consumption pc (log) *Landless*	Consumption pc (log) *Landowners*
Below canal	-0.025***	-0.053***	-0.004	0.003	0.021***
	(0.005)	(0.018)	(0.014)	(0.006)	(0.005)
Control group mean	0.536	0.750	1.502	9.604	9.811
Observations	86,117	83,796	83,763	83,802	84,108
R^2	0.460	0.460	0.510	0.460	0.560

Panel B. Consumption distribution

	Consumption pc (log) 1st quintile *Landowners*	Consumption pc (log) 2nd quintile *Landowners*	Consumption pc (log) 3rd quintile *Landowners*	Consumption pc (log) 4th quintile *Landowners*	Consumption pc (log) 5th quintile *Landowners*
Below canal	0.001	0.015**	0.014**	0.022***	0.031***
	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)
Control group mean	9.739	9.766	9.805	9.846	9.934
Observations	74,140	76,446	72,323	74,105	68,911
R^2	0.460	0.460	0.410	0.410	0.380

Panel C. Education attainment

	At least Primary (share of adult pop.)		At least Middle (share of adult pop.)		At least Secondary (share of adult pop.)	
	Landless	*Landowners*	*Landless*	*Landowners*	*Landless*	*Landowners*
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: Regression discontinuity results following Equation 5.1 separately estimated for landowning and landless households. Panel A summarizes differences in the share of landowners, the size of landowners' plots, and the overall consumption of landowners and the landless in each settlement. Panel B shows the consumption of landowners by quintile of land holding size. The bottom (1st) quintile are the landowners with plots in the 0-20% range of the national distribution while the top (5th) quintile are those landowners with total land holdings in the top 80-100% of the national distribution. The quintile break points in ascending order are 1.32, 2.30, 3.65, and 6.10 acres. Note that all consumption coefficients are in units of log consumption per capita, as htey 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	0.074***	0.093***	0.011	-0.018*
	(0.016)	(0.011)	(0.008)	(0.009)
Above canal	0.016*	0.007*	0.018**	-0.009*
	(0.009)	(0.004)	(0.008)	(0.005)
Control group mean	0.381	0.034	0.196	0.156
Observations	80,408	80,572	80,576	79,864
R^2	0.60	0.20	0.39	0.77

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water crops (any)
Below canal	0.018**	0.011	0.016	0.055**
	(0.008)	(0.016)	(0.021)	(0.022)
Above canal	-0.002	0.007	-0.029	0.039**
	(0.008)	(0.013)	(0.019)	(0.017)
Control group mean	0.569	7.808	7.329	0.632
Observations	90,055	89,997	89,800	69,287
R^2	0.55	0.86	0.56	0.70

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 households)
Below canal	0.191***	0.011	0.006	0.005	0.011
	(0.030)	(0.007)	(0.004)	(0.004)	(0.009)
Above canal	0.038*	0.006	0.003	0.003	-0.008
	(0.023)	(0.007)	(0.003)	(0.004)	(0.008)
Control group mean	5.524	0.109	0.069	0.032	9.636
Observations	91,267	83,986	83,986	83,986	86,640
R^2	0.29	0.01	0.00	0.02	0.43

Panel D. Outcomes disaggregated by landownership

	Consumption pc (log) *Landless*	Consumption pc (log) *Landowners*	Middle school ed. *Landless*	Middle school ed. *Landowners*
Below canal	-0.013	0.018*	0.007*	0.026***
	(0.008)	(0.009)	(0.004)	(0.006)
Above canal	-0.015*	-0.002	0.000	0.004
	(0.008)	(0.009)	(0.004)	(0.005)
Control group mean	9.514	9.739	0.252	0.335
Observations	83,497	84,052	83,303	84,012
R^2	0.38	0.43	0.44	0.53

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

Notes: This table shows the results of the spillovers analysis, comparing above- and below-canal settlements to distant settlements far from the canal within the same district, following Equation 5.2. 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 coefficients on the dummy variables for being in the below-canal treatment or above-canal control groups are reported here.

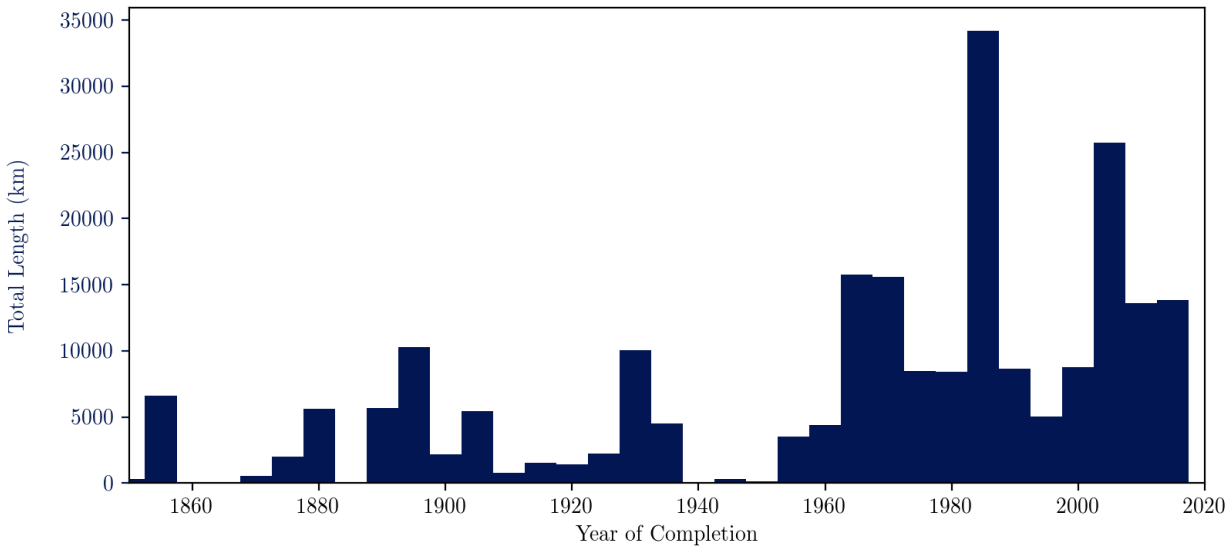
Table 6: Effect of canals on urbanization (event study)

	Town Existence (pop. 5,000)		Population (log)		Growth (decadal)	
	1	2	3	4	5	6
Command area in town catchment area	0.039**		0.084***		0.027	
(binary treatment)	(0.016)		(0.031)		(0.024)	
Share of town catchment area in command area		0.054*		0.334***		0.074**
(continuous treatment)		(0.032)		(0.084)		(0.032)
Observations	26,292	71,520	26,292	71,520	24,101	65,560
R^2		0.70		0.82		0.15

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

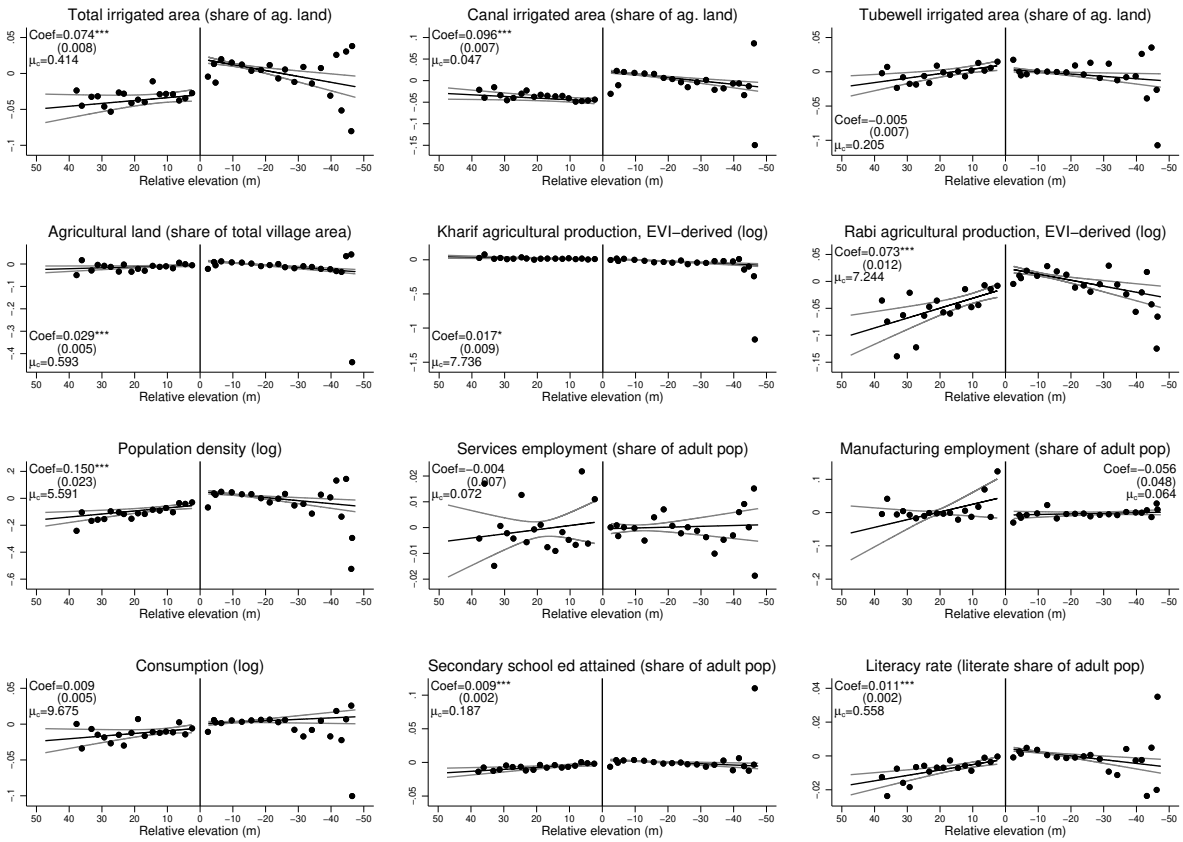
Notes: This table shows the estimated effect of canal area on town growth as estimated in Equation 5.3. Each column reports the β_1 values for various outcome variables. The outcome variable in columns 1 and 2 is the existence of a town with population 5,000 or greater as the outcome variable. In columns 3 and 4 the outcome variable is log population while in columns 5 and 6 it is decadal population growth. All regressions assume that before a town appears in the time series, its is a settlements with a population of 2,000 (smaller than the population required to be declared a town). The first row shows Callaway-Sant'anna results using a binary indicator for canal construction where a town is treated when 20% of catchment area (radius=20km) around it has been covered by a command area. The second row uses a continuous value, the share of the town catchment area (radius=20km) covered by a command area, as the dependent variable.

Figure 1: Canal construction through time



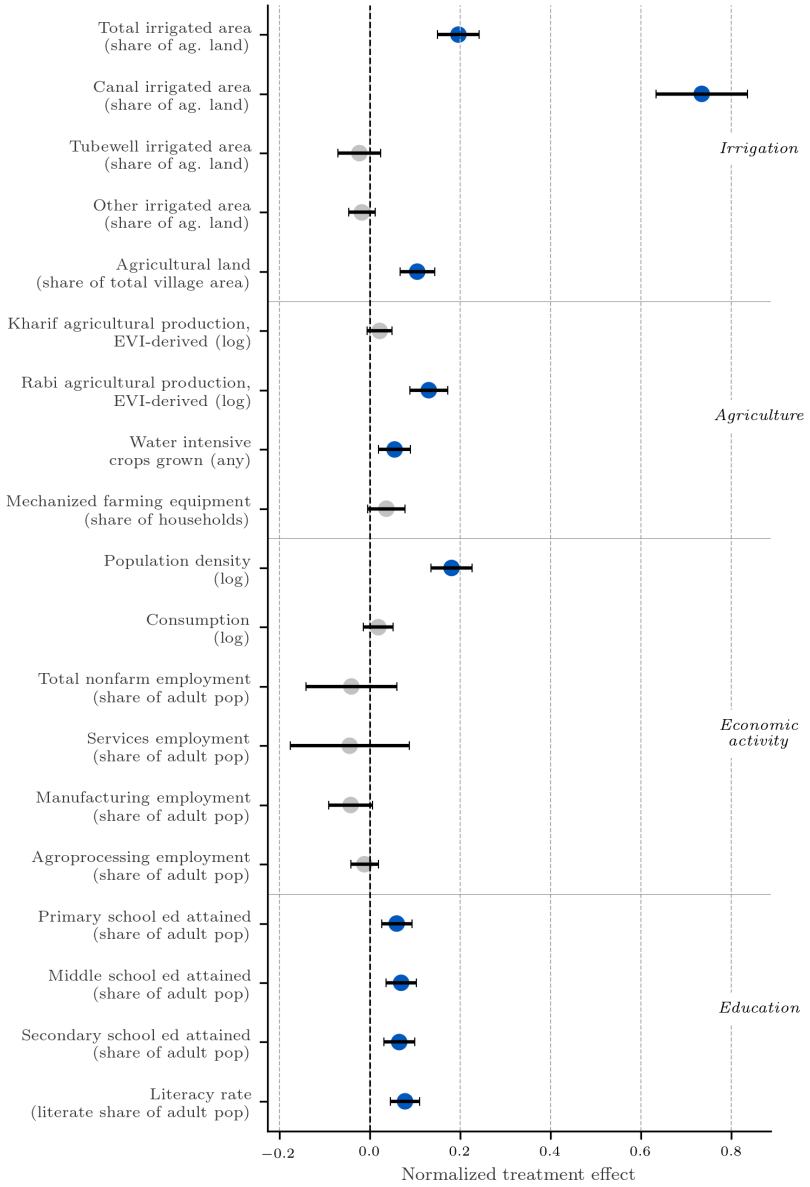
Notes: The total length of medium and major canals constructed in India from 1850-2013. Any canals with dates older than 1850 are coded as 1850 while any canals not completed before 2013 are not included. Note that 217 of the 1442 total canal projects reported, or 9% of total canal length in the geospatial 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 outcome variables) and so are not included in this plot.

Figure 2: Regression discontinuity binscatters for key variables



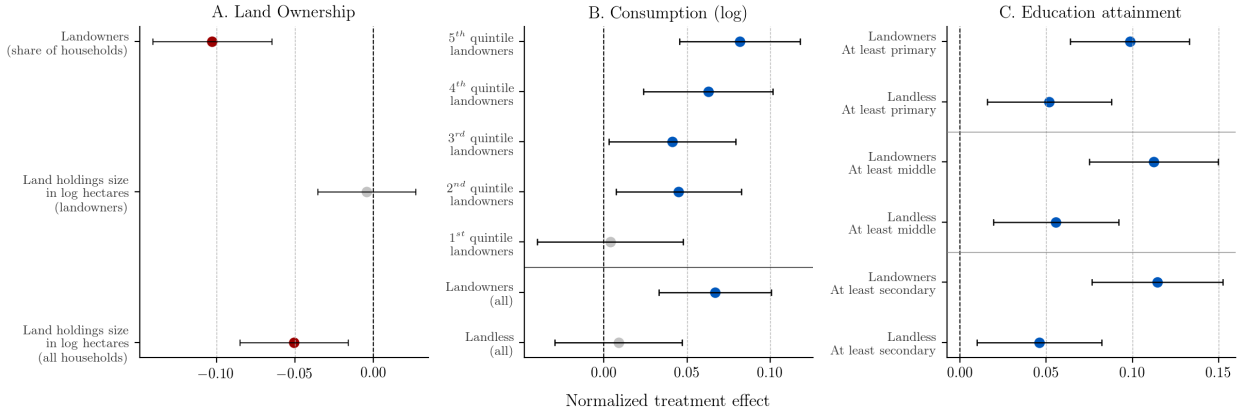
Notes: Each figure shows the binned scatterplot relationship between an outcome of interest and the RDD running variable (elevation relative to the nearest canal), after residualizing on the geophysical controls and subdistrict fixed effects. Treatment settlements that lie below the canal have negative relative elevation while untreated settlements that lie above the canal have positive relative elevation. All regressions were done following 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 coefficients for main outcomes



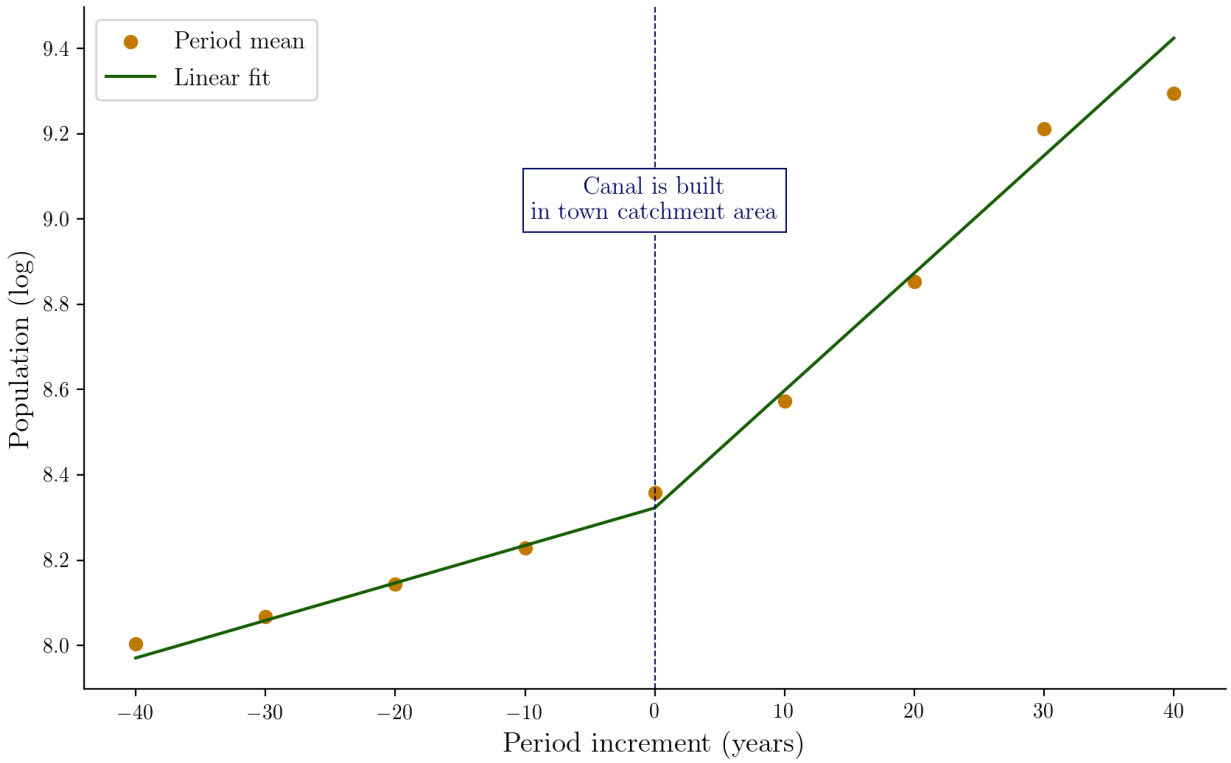
Notes: This figure shows the regression discontinuity results for our main outcomes using our preferred specification and following Equation 5.1. Blue points indicate normalized treatment effects above 0, red below zero, and gray indicates an insignificant result. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in the sample. Error bars indicate the 95% confidence interval for each estimate.

Figure 4: Land ownership outcomes



Notes: This figure shows the regression discontinuity coefficients for outcomes disaggregated by land ownership. Following Equation 5.1, we estimate the RDD coefficients for outcomes based on land ownership status. Blue points indicate normalized treatment effects above 0, red below zero, and gray indicates an insignificant result. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in the 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, with the 20 km town radius used in this figure. Period increment 10 indicates 1 decade after the command area appearance while period increment -10 indicates 1 decade before the first command area appearance.

Table A1: Additional RDD 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.)
Panel A: Relative elevation				
Below canal	0.009	-0.002***	0.000	-0.024
	(0.007)	(0.001)	(0.000)	(0.024)
Control group mean	0.025	0.140	0.037	0.018
Observations	91,465	91,465	86,111	85,342
R^2	0.150	0.570	0.340	0.040

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

Notes: Additional outcomes reported for the relative elevation specification of the regression discontinuity design. These results use the balanced, analysis sample following Equation 5.1.

Table A2: Balance in the RDD using distance to command area boundary

	Ruggedness	Annual rainfall avg. 2010-2014 (mm)	Distance to coast (km)	Distance to river (km)	Wetland rice (GAEZ)	Wheat (GAEZ)
Below canal	-0.022	-1.130	-0.062	-0.843	-0.008	0.009
	(0.040)	(3.139)	(0.459)	(0.769)	(0.016)	(0.006)
Control group mean	3.602	1317.887	467.054	23.404	2.704	1.187
Observations	48,965	48,965	48,965	48,965	48,965	48,965
R^2	0.640	0.990	1.000	0.930	0.960	0.990

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

Notes: This table reports the regression discontinuity effect on several outcomes we expect to be balanced using the command area boundary RDD, which uses distance to the command area boundary as the running variable instead of relative elevation. Crop suitability measures are taken from the Global Agro-Ecological Zones model that estimates expected conditions for agricultural production based on climate, soil, and terrain parameters. Model estimates assume that crops have gravity-fed irrigation and intermediate inputs.

Table A3: Regression discontinuity irrigation outcomes for a range of 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	-0.008***
	(0.002)	(0.001)	(0.000)	(0.000)
Control group mean	0.435	0.070	0.214	0.158
Observations	230,139	230,190	230,282	228,312
R^2	0.640	0.450	0.470	0.560
Panel B: All canal-area settlements, minus donut hole				
Below canal	0.079***	0.107***	-0.007	-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
	(0.003)	(0.004)	(0.000)	(-0.001)
Control group mean	0.469	0.068	0.245	0.162
Observations	90,434	90,435	90,496	89,492
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***	0.103***	-0.014***	-0.007*
	(0.002)	(0.003)	(0.000)	(0.000)
Control group mean	0.440	0.055	0.226	0.166
Observations	94,646	94,656	94,713	93,726
R^2	0.650	0.450	0.500	0.620
Panel E: Command area boundary RDD				
Inside command area	0.118***	0.156***	-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
Panel F: Main analysis results				
Below canal	0.074***	0.096***	-0.005	-0.007
	(0.002)	(0.002)	(0.000)	(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	0.390	0.470	0.630

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

Notes: This table shows irrigation outcomes using a range of samples to test the robustness of our main results of the regression discontinuity. Panel A contains all settlements $\leq$10km from the nearest canal in distance and $\pm$50m from the nearest canal in elevation. Panel B removes the "donut hole", meaning all settlements $\pm$2.5m from the canal are removed. The samples in Panel C and D use the same sample restrictions as as our main, balanced analysis sample (shown for reference in panel F), but use different values for settlement elevation. Panel C uses the median while Panel D uses the 25th percentile to parameterize settlement elevation when calculating relative elevation. Panel E shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A4: Regression discontinuity results for agricultural outcomes for all 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 households)
Panel A: All canal-area settlements					
Below canal	0.032***	0.028***	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	242,163	242,220	240,423	193,861	231,571
R^2	0.590	0.790	0.720	0.730	0.340
Panel B: All canal-area settlements, minus donut hole					
Below canal	0.042***	0.026***	0.066***	0.029***	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	0.630	7.747	7.292	0.691	0.049
Observations	97,507	97,440	97,077	77,316	93,412
R^2	0.610	0.820	0.730	0.730	0.350
Panel D: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation					
Below canal	0.029***	0.033***	0.060***	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	101,919	101,876	101,595	80,408	97,503
R^2	0.620	0.820	0.730	0.720	0.340
Panel E: Command area boundary RDD					
Inside command area	0.026**	0.141***	0.061**	0.028	0.008**
	(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
Panel F: Main analysis results					
Below canal	0.029***	0.017*	0.073***	0.028***	0.002
	(0.002)	(0.003)	(0.003)	(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

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

Notes: This table shows agricultural outcomes using a range of samples to test the robustness of our main results of the regression discontinuity. Panel A contains all settlements ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation. Panel B removes the "donut hole", meaning all settlements ±2.5m from the canal are removed. The samples in Panel C and D use the same sample restrictions as as our main, balanced analysis sample (shown for reference in panel F), but use different values for settlement elevation. Panel C uses the median while Panel D uses the 25th percentile to parameterize settlement elevation when calculating relative elevation. Panel E shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A5: Regression discontinuity results for non-farm outcomes for all 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*	Consumption pc (log) *Landowners*
Panel A: All canal-area settlements						
Below canal	0.123***	-0.027	-0.002	-0.024	0.004	0.019***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.692	0.142	0.078	0.048	9.535	9.735
Observations	245,531	226,911	226,911	226,911	225,320	225,921
R^2	0.510	0.020	0.010	0.020	0.470	0.590
Panel B: All canal-area settlements, minus donut hole						
Below canal	0.185***	-0.041	-0.002	-0.037	0.009*	0.026***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.527	0.146	0.076	0.052	9.536	9.726
Observations	132,969	123,496	123,496	123,496	122,037	122,044
R^2	0.450	0.020	0.010	0.020	0.480	0.550
Panel C: Canal-area settlements balanced on ruggedness, using median settlement elevation						
Below canal	0.029*	0.004	-0.003	0.000	-0.001	0.008*
	(0.007)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Control group mean	5.802	0.149	0.076	0.061	9.542	9.757
Observations	99,102	92,493	92,493	92,493	91,030	91,235
R^2	0.490	0.020	0.020	0.020	0.500	0.600
Panel D: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation						
Below canal	0.109***	-0.018	-0.002	-0.018	0.004	0.021***
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Control group mean	5.696	0.145	0.075	0.056	9.550	9.756
Observations	103,548	96,668	96,668	96,668	94,966	95,298
R^2	0.500	0.020	0.020	0.020	0.480	0.590
Panel E: Command area boundary RDD						
Inside command area	0.226***	0.005	-0.004	0.004	0.000	0.034***
	(0.002)	(0.000)	(0.000)	(0.000)	(-0.001)	(0.000)
Control group mean	6.281	0.114	0.080	0.029	9.516	9.770
Observations	48,909	45,098	45,098	45,098	45,029	44,837
R^2	0.650	0.050	0.050	0.030	0.500	0.580
Panel F: Main analysis results						
Below canal	0.150***	-0.063	-0.004	-0.056	0.003	0.021***
	(0.003)	(-0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	5.590	0.151	0.072	0.064	9.550	9.747
Observations	91,465	85,342	85,342	85,342	83,802	84,108
R^2	0.460	0.020	0.010	0.020	0.470	0.560

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

Notes: This table shows non-farm outcomes using a range of samples to test the robustness of our main results of the regression discontinuity. Panel A contains all settlements $\leq$10km from the nearest canal in distance and $\pm$50m from the nearest canal in elevation. Panel B removes the "donut hole", meaning all settlements $\pm$2.5m from the canal are removed. The samples in Panel C and D use the same sample restrictions as as our main, balanced analysis sample (shown for reference in panel F), but use different values for settlement elevation. Panel C uses the median while Panel D uses the 25th percentile to parameterize settlement elevation when calculating relative elevation. Panel E shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A6: Regression discontinuity results for education outcomes for all specifications

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (share of pop.)
Panel A: All canal-area settlement				
Below canal	0.015***	0.015***	0.011***	0.009***
	(0.001)	(0.001)	(0.000)	(0.000)
Control group mean	0.462	0.308	0.185	0.554
Observations	231,452	231,452	231,452	245,531
R^2	0.580	0.560	0.530	0.600
Panel B: All canal-area settlements, minus donut hole				
Below canal	0.021***	0.020***	0.015***	0.013***
	(0.001)	(0.000)	(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	93,374	93,374	93,374	99,102
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***	0.010***	0.009***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Control group mean	0.475	0.317	0.192	0.564
Observations	97,460	97,460	97,460	103,548
R^2	0.580	0.570	0.550	0.610
Panel E: 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
Panel F: Main analysis results				
Below canal	0.013***	0.012***	0.009***	0.011***
	(0.001)	(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

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

Notes: This table shows education outcomes using a range of samples to test the robustness of our main results of the regression discontinuity. Panel A contains all settlements $\leq$10km from the nearest canal in distance and $\pm$50m from the nearest canal in elevation. Panel B removes the "donut hole", meaning all settlements $\pm$2.5m from the canal are removed. The samples in Panel C and D use the same sample restrictions as as our main, balanced analysis sample (shown for reference in panel F), but use different values for settlement elevation. Panel C uses the median while Panel D uses the 25th percentile to parameterize settlement elevation when calculating relative elevation. Panel E shows the results of the regression discontinuity using the distance to command area boundary specification.

Table A7: RD analysis sensitivity to bandwidth and canal distance

Panel A. Regression discontinuity bandwidth

Bandwidth (m)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
25	0.065***	0.054***	0.106***	0.009***	-0.044	88,535
	(0.009)	(0.012)	(0.024)	(0.002)	(0.057)	
50	0.068***	0.080***	0.150***	0.011***	-0.008	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.056)	
75	0.075***	0.071***	0.165***	0.010***	-0.056	90,735
	(0.008)	(0.012)	(0.023)	(0.002)	(0.055)	

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%	0.065***	0.046***	0.126***	0.006**	-0.011	54,914
	(0.011)	(0.016)	(0.028)	(0.003)	(0.037)	
25%	0.074***	0.073***	0.150***	0.011***	-0.008	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.056)	
50%	0.075***	0.068***	0.167***	0.011***	-0.167***	116,695
	(0.007)	(0.011)	(0.019)	(0.002)	(0.053)	

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	0.082***	0.062***	0.179***	0.008***	-0.015	61,217
	(0.012)	(0.016)	(0.028)	(0.003)	(0.052)	
10	0.074***	0.073***	0.150***	0.011***	-0.008	91,465
	(0.008)	(0.012)	(0.023)	(0.002)	(0.056)	
15	0.069***	0.077***	0.143***	0.003	-0.045	109,071
	(0.007)	(0.011)	(0.020)	(0.007)	(0.044)	

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

Notes: This table shows the sensitivity of our major results to changes in the construction of our sample. Each panel shows the results if one assumption were to be changed. The bolded parameters indicate the values used in our preferred 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 treatment and control settlements for each fixed effect group. 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: Spillovers analysis robustness for irrigation outcomes using entropy balance

	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	0.070***	0.093***	0.006	-0.018**
	(0.016)	(0.011)	(0.007)	(0.009)
Above canal	0.012	0.005	0.014*	-0.006
	(0.009)	(0.004)	(0.008)	(0.005)
Control group mean	0.383	0.037	0.189	0.162
Observations	104,083	104,268	104,265	103,469
R^2	0.60	0.19	0.39	0.76
Panel B. Entropy balance, 1% outliers dropped				
Below canal	0.072***	0.095***	0.007	-0.019**
	(0.016)	(0.011)	(0.008)	(0.009)
Above canal	0.016*	0.007*	0.017**	-0.007
	(0.009)	(0.004)	(0.008)	(0.005)
Control group mean	0.380	0.036	0.193	0.155
Observations	93,834	94,012	94,014	93,246
R^2	0.59	0.19	0.39	0.76
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**				
Below canal	0.074***	0.093***	0.011	-0.018*
	(0.016)	(0.011)	(0.008)	(0.009)
Above canal	0.016*	0.007*	0.018**	-0.009*
	(0.009)	(0.004)	(0.008)	(0.005)
Control group mean	0.381	0.034	0.196	0.156
Observations	80,408	80,572	80,576	79,864
R^2	0.60	0.20	0.39	0.77
Panel D. Entropy balance, 5% outliers dropped				
Below canal	0.069***	0.103***	0.007	-0.028**
	(0.019)	(0.013)	(0.010)	(0.011)
Above canal	0.008	0.006	0.012	-0.010
	(0.010)	(0.005)	(0.009)	(0.007)
Control group mean	0.380	0.029	0.205	0.150
Observations	63,864	64,027	64,020	63,385
R^2	0.63	0.20	0.41	0.77

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all irrigation outcomes. Results follow Equation 5.2 using entropy balancing method as described in the main results. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main text), and Panel D drops 5% outliers.

Table A9: Spillovers analysis robustness for agricultre outcomes using entropy balance

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water crops (any)
Panel A. Entropy balance, no outliers dropped				
Below canal	0.017*	-0.002	0.043*	0.068***
	(0.009)	(0.019)	(0.022)	(0.022)
Above canal	-0.005	0.009	-0.021	0.037**
	(0.009)	(0.016)	(0.018)	(0.017)
Control group mean	0.562	7.760	7.329	0.648
Observations	115,251	115,413	115,158	88,658
R^2	0.53	0.83	0.57	0.65
Panel B. Entropy balance, 1% outliers dropped				
Below canal	0.022***	0.006	0.037*	0.069***
	(0.008)	(0.020)	(0.021)	(0.021)
Above canal	0.002	0.011	-0.020	0.046***
	(0.008)	(0.017)	(0.018)	(0.017)
Control group mean	0.565	7.784	7.325	0.642
Observations	104,415	104,384	104,139	80,371
R^2	0.54	0.84	0.56	0.68
Panel C. Entropy balance, 2.5% outliers dropped				
Below canal	0.018**	0.011	0.016	0.055**
	(0.008)	(0.016)	(0.021)	(0.022)
Above canal	-0.002	0.007	-0.029	0.039**
	(0.008)	(0.013)	(0.019)	(0.017)
Control group mean	0.569	7.808	7.329	0.632
Observations	90,055	89,997	89,800	69,287
R^2	0.55	0.86	0.56	0.70
Panel D. Entropy balance, 5% outliers dropped				
Below canal	0.018**	0.018	-0.017	0.050**
	(0.009)	(0.015)	(0.023)	(0.024)
Above canal	-0.004	0.000	-0.040*	0.048**
	(0.008)	(0.012)	(0.021)	(0.023)
Control group mean	0.584	7.826	7.368	0.613
Observations	71,244	71,202	71,069	53,857
R^2	0.56	0.88	0.57	0.70

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all agricultre outcomes. Results follow Equation 5.2 using entropy balancing method as described in the main results. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main text), and Panel D drops 5% outliers.

Table A10: Spillovers analysis robustness for non-farm outcomes using entropy balance

	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 households)
Panel A. Entropy balance, no outliers dropped					
Below canal	0.182***	0.003	0.004	0.000	0.012
	(0.031)	(0.006)	(0.003)	(0.004)	(0.008)
Above canal	0.016	-0.004	0.000	-0.003	-0.014*
	(0.026)	(0.008)	(0.003)	(0.006)	(0.008)
Control group mean	5.528	0.127	0.080	0.034	9.634
Observations	117,083	107,401	107,401	107,401	111,400
R^2	0.32	0.01	0.00	0.01	0.41
Panel B. Entropy balance, 1% outliers dropped					
Below canal	0.198***	0.006	0.004	0.003	0.014*
	(0.027)	(0.005)	(0.003)	(0.003)	(0.008)
Above canal	0.041*	0.002	0.001	0.002	-0.008
	(0.023)	(0.006)	(0.003)	(0.004)	(0.007)
Control group mean	5.534	0.120	0.077	0.033	9.634
Observations	105,874	97,379	97,379	97,379	100,701
R^2	0.29	0.01	0.00	0.02	0.42
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**					
Below canal	0.191***	0.011	0.006	0.005	0.011
	(0.030)	(0.007)	(0.004)	(0.004)	(0.009)
Above canal	0.038*	0.006	0.003	0.003	-0.008
	(0.023)	(0.007)	(0.003)	(0.004)	(0.008)
Control group mean	5.524	0.109	0.069	0.032	9.636
Observations	91,267	83,986	83,986	83,986	86,640
R^2	0.29	0.01	0.00	0.02	0.43
Panel D. Entropy balance, 5% outliers dropped					
Below canal	0.210***	0.011	0.008*	0.004	0.014
	(0.035)	(0.007)	(0.004)	(0.004)	(0.011)
Above canal	0.026	0.008	0.004	0.004	-0.010
	(0.025)	(0.007)	(0.004)	(0.004)	(0.009)
Control group mean	5.492	0.102	0.065	0.029	9.641
Observations	72,192	66,484	66,484	66,484	68,366
R^2	0.28	0.01	0.00	0.03	0.40

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all non-farm outcomes. Results follow Equation 5.2 using entropy balancing method as described in the main results. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main text), and Panel D drops 5% outliers.

Table A11: Spillovers analysis robustness for outcomes disaggregated by landownership using entropy balance

	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 households)
Panel A. Entropy balance, no outliers dropped					
Below canal	0.182***	0.003	0.004	0.000	0.012
	(0.031)	(0.006)	(0.003)	(0.004)	(0.008)
Above canal	0.016	-0.004	0.000	-0.003	-0.014*
	(0.026)	(0.008)	(0.003)	(0.006)	(0.008)
Control group mean	5.528	0.127	0.080	0.034	9.634
Observations	117,083	107,401	107,401	107,401	111,400
R^2	0.32	0.01	0.00	0.01	0.41
Panel B. Entropy balance, 1% outliers dropped					
Below canal	0.198***	0.006	0.004	0.003	0.014*
	(0.027)	(0.005)	(0.003)	(0.003)	(0.008)
Above canal	0.041*	0.002	0.001	0.002	-0.008
	(0.023)	(0.006)	(0.003)	(0.004)	(0.007)
Control group mean	5.534	0.120	0.077	0.033	9.634
Observations	105,874	97,379	97,379	97,379	100,701
R^2	0.29	0.01	0.00	0.02	0.42
Panel C. Entropy balance, 2.5% outliers dropped - **preferred specification**					
Below canal	0.191***	0.011	0.006	0.005	0.011
	(0.030)	(0.007)	(0.004)	(0.004)	(0.009)
Above canal	0.038*	0.006	0.003	0.003	-0.008
	(0.023)	(0.007)	(0.003)	(0.004)	(0.008)
Control group mean	5.524	0.109	0.069	0.032	9.636
Observations	91,267	83,986	83,986	83,986	86,640
R^2	0.29	0.01	0.00	0.02	0.43
Panel D. Entropy balance, 5% outliers dropped					
Below canal	0.210***	0.011	0.008*	0.004	0.014
	(0.035)	(0.007)	(0.004)	(0.004)	(0.011)
Above canal	0.026	0.008	0.004	0.004	-0.010
	(0.025)	(0.007)	(0.004)	(0.004)	(0.009)
Control group mean	5.492	0.102	0.065	0.029	9.641
Observations	72,192	66,484	66,484	66,484	68,366
R^2	0.28	0.01	0.00	0.03	0.40

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for outcomes disaggregated by landownership. Results follow Equation 5.2 using entropy balancing method as described in the main results. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5% (as in the main text), and Panel D drops 5% outliers.

Table A12: Spillovers analysis robustness for irrigation outcomes using coarsened exact matching

	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	0.056***	0.073***	0.002	-0.015*
	(0.013)	(0.009)	(0.008)	(0.009)
Above canal	0.000	0.004	0.005	-0.010
	(0.009)	(0.005)	(0.007)	(0.007)
Control group mean	0.415	0.041	0.209	0.171
Observations	47,144	47,262	47,240	46,841
R^2	0.55	0.21	0.42	0.66
Panel B. Coarsened exact matching, 1% outliers dropped				
Below canal	0.054***	0.066***	0.002	-0.009
	(0.014)	(0.009)	(0.009)	(0.010)
Above canal	-0.003	0.004	0.002	-0.009
	(0.010)	(0.005)	(0.008)	(0.008)
Control group mean	0.441	0.031	0.228	0.186
Observations	32,185	32,235	32,231	31,981
R^2	0.57	0.21	0.43	0.69
Panel C. Coarsened exact matching, 2.5% outliers dropped				
Below canal	0.056***	0.073***	-0.002	-0.011
	(0.014)	(0.011)	(0.010)	(0.009)
Above canal	0.003	0.009	0.006	-0.010
	(0.011)	(0.006)	(0.009)	(0.008)
Control group mean	0.436	0.032	0.224	0.188
Observations	21,369	21,398	21,388	21,240
R^2	0.58	0.24	0.44	0.68
Panel D. Coarsened exact matching, 5% outliers dropped				
Below canal	0.037**	0.071***	-0.009	-0.022**
	(0.017)	(0.012)	(0.008)	(0.011)
Above canal	-0.002	0.006	0.004	-0.016*
	(0.012)	(0.006)	(0.010)	(0.009)
Control group mean	0.436	0.020	0.223	0.197
Observations	13,147	13,163	13,160	13,071
R^2	0.58	0.22	0.45	0.70

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all irrigation outcomes. Results follow Equation 5.2 employing coarsened exact matching (CEM), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers.

Table A13: Spillovers analysis robustness for agriculture outcomes using coarsened exact matching

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water crops (any)
Panel A. Coarsened exact matching, no outliers dropped				
Below canal	0.015*	-0.040***	0.011	0.028*
	(0.008)	(0.012)	(0.023)	(0.015)
Above canal	0.005	-0.005	0.001	0.013
	(0.008)	(0.010)	(0.019)	(0.013)
Control group mean	0.602	7.823	7.345	0.616
Observations	53,111	53,074	52,936	40,881
R^2	0.55	0.75	0.59	0.74
Panel B. Coarsened exact matching, 1% outliers dropped				
Below canal	0.009	-0.029**	0.010	0.025
	(0.010)	(0.014)	(0.022)	(0.015)
Above canal	0.005	-0.002	0.002	0.019
	(0.010)	(0.013)	(0.021)	(0.014)
Control group mean	0.645	7.868	7.434	0.584
Observations	36,432	36,402	36,327	27,819
R^2	0.60	0.77	0.62	0.72
Panel C. Coarsened exact matching, 2.5% outliers dropped				
Below canal	0.014	-0.033**	0.019	0.059***
	(0.009)	(0.014)	(0.025)	(0.020)
Above canal	0.008	-0.006	0.011	0.048**
	(0.009)	(0.012)	(0.024)	(0.021)
Control group mean	0.662	7.898	7.397	0.563
Observations	24,115	24,105	24,042	18,269
R^2	0.62	0.79	0.64	0.71
Panel D. Coarsened exact matching, 5% outliers dropped				
Below canal	0.006	-0.034**	0.028	0.033
	(0.010)	(0.016)	(0.024)	(0.022)
Above canal	0.005	-0.001	0.024	0.023
	(0.011)	(0.015)	(0.019)	(0.022)
Control group mean	0.650	7.935	7.400	0.551
Observations	14,644	14,646	14,595	10,998
R^2	0.64	0.79	0.66	0.71

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all agriculture outcomes. Results follow Equation 5.2 employing coarsened exact matching (CEM), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers.

Table A14: Spillovers analysis robustness for non-farm outcomes using coarsened exact matching

	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 households)
Panel A. Coarsened exact matching, no outliers dropped					
Below canal	0.115***	0.004	0.008	-0.001	0.018**
	(0.027)	(0.010)	(0.008)	(0.003)	(0.009)
Above canal	0.047**	0.004	0.001	0.005	0.004
	(0.022)	(0.015)	(0.013)	(0.003)	(0.008)
Control group mean	5.546	0.119	0.079	0.033	9.661
Observations	53,865	50,066	50,066	50,066	51,264
R^2	0.32	0.00	0.00	0.02	0.39
Panel B. Coarsened exact matching, 1% outliers dropped					
Below canal	0.088**	0.006	0.011	-0.001	0.005
	(0.034)	(0.012)	(0.010)	(0.004)	(0.009)
Above canal	0.049	0.010	0.008	0.005	-0.003
	(0.031)	(0.017)	(0.015)	(0.004)	(0.009)
Control group mean	5.535	0.102	0.067	0.027	9.704
Observations	36,886	34,448	34,448	34,448	35,274
R^2	0.34	0.00	0.00	0.03	0.39
Panel C. Coarsened exact matching, 2.5% outliers dropped					
Below canal	0.098***	0.004	0.009	-0.001	0.004
	(0.029)	(0.012)	(0.009)	(0.006)	(0.010)
Above canal	0.048*	0.006	0.006	0.004	0.000
	(0.026)	(0.014)	(0.012)	(0.005)	(0.010)
Control group mean	5.508	0.129	0.086	0.035	9.695
Observations	24,419	22,811	22,811	22,811	23,299
R^2	0.33	0.00	0.00	0.04	0.38
Panel D. Coarsened exact matching, 5% outliers dropped					
Below canal	0.084***	0.007	0.008	0.003	-0.007
	(0.031)	(0.014)	(0.012)	(0.006)	(0.010)
Above canal	0.046*	0.008	0.008	0.004	-0.010
	(0.026)	(0.013)	(0.011)	(0.005)	(0.010)
Control group mean	5.446	0.088	0.062	0.021	9.690
Observations	14,822	13,848	13,848	13,848	14,243
R^2	0.36	-0.01	-0.01	-0.01	0.37

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for all non-farm outcomes. Results follow Equation 5.2 employing coarsened exact matching (CEM), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers.

Table A15: Spillovers analysis robustness for outcomes disaggregated by landownership using coarsened exact matching

	Consumption (log) *Landless*	Consumption (log) *Landowners*	Middle school ed. *Landless*	Middle school ed. *Landowners*
Panel A. Coarsened exact matching, no outliers dropped				
Below canal	-0.006	0.033***	0.017***	0.040***
	(0.008)	(0.010)	(0.005)	(0.007)
Above canal	-0.011	0.010	0.008*	0.016***
	(0.009)	(0.010)	(0.005)	(0.006)
Control group mean	9.533	9.765	0.259	0.345
Observations	49,421	49,792	49,319	49,765
R^2	0.32	0.40	0.35	0.48
Panel B. Coarsened exact matching, 1% outliers dropped				
Below canal	-0.014	0.021**	0.010**	0.031***
	(0.011)	(0.010)	(0.005)	(0.006)
Above canal	-0.016	0.004	0.007	0.012**
	(0.010)	(0.010)	(0.005)	(0.005)
Control group mean	9.556	9.809	0.266	0.362
Observations	34,154	34,336	34,094	34,319
R^2	0.33	0.40	0.38	0.48
Panel C. Coarsened exact matching, 2.5% outliers dropped				
Below canal	-0.017*	0.029**	0.019***	0.034***
	(0.010)	(0.012)	(0.006)	(0.008)
Above canal	-0.021*	0.012	0.015**	0.014**
	(0.011)	(0.012)	(0.006)	(0.007)
Control group mean	9.547	9.790	0.257	0.350
Observations	22,573	22,734	22,531	22,721
R^2	0.31	0.38	0.36	0.47
Panel D. Coarsened exact matching, 5% outliers dropped				
Below canal	-0.006	0.007	0.019***	0.035***
	(0.011)	(0.012)	(0.007)	(0.008)
Above canal	-0.010	-0.004	0.013*	0.021***
	(0.011)	(0.011)	(0.007)	(0.008)
Control group mean	9.547	9.786	0.256	0.343
Observations	13,775	13,883	13,741	13,877
R^2	0.32	0.36	0.37	0.47

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

Notes: This table shows robustness to the spillovers analysis results presented in Table 5 for outcomes disaggregated by landownership. Results follow Equation 5.2 employing coarsened exact matching (CEM), which discretizes continuous control variables into bins before drawing balanced groups from across the coarsened distributions. Panel A does not drop any outliers while Panel B drops 1%, Panel C drops 2.5%, and Panel C drops 5% outliers.

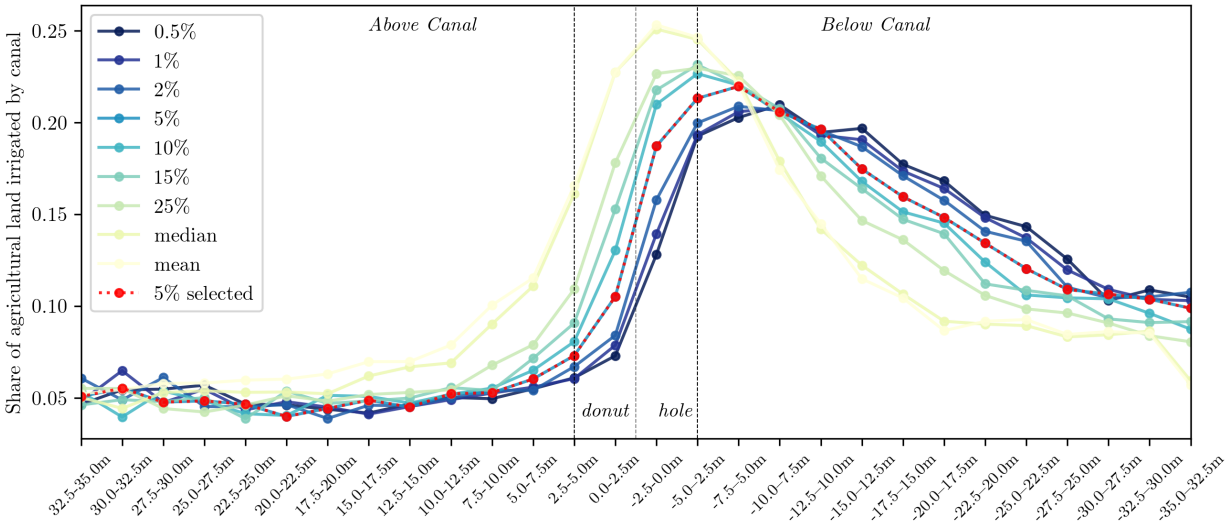
Table A16: Town analysis robustness

	Town Existence (pop. 5,000)			Population (log)			Growth (decadal)		
Panel A. 10km radius									
	1	2	3	4	5	6	7	8	9
Command area in town catchment area	0.030*			0.065**			0.057***		
(binary treatment)	(0.017)			(0.031)			(0.021)		
Share of 0-10km band in command area		0.056*	0.117		0.273***	0.241		0.064**	0.084
(continuous treatment)		(0.032)	(0.079)		(0.070)	(0.151)		(0.025)	(0.060)
Share of 10-20km band in command area			-0.078			0.041			-0.026
			(0.076)			(0.163)			(0.065)
Observations	25,416	74,952	74,952	25,416	74,952	74,952	23,298	68,706	68,706
R^2		0.70	0.67		0.82	0.80		0.15	0.06
Panel B. 20km radius									
	1	2	3	4	5	6	7	8	9
Command area in town catchment area	0.039**			0.084***			0.027		
(binary treatment)	(0.016)			(0.031)			(0.024)		
Share of 0-20km band in command area		0.054*	0.139*		0.334***	0.284*		0.074**	0.202***
(continuous treatment)		(0.032)	(0.079)		(0.084)	(0.171)		(0.032)	(0.070)
Share of 20-40km band in command area			-0.121			0.070			-0.182**
			(0.084)			(0.210)			(0.074)
Observations	26,292	71,520	71,520	26,292	71,520	71,520	24,101	65,560	65,560
R^2		0.70	0.67		0.82	0.81		0.15	0.06
Panel C. 30km radius									
	1	2	3	4	5	6	7	8	9
Command area in town catchment area	0.045***			0.098***			0.037		
(binary treatment)	(0.017)			(0.031)			(0.023)		
Share of 0-30km band in command area		0.056*	0.088		0.390***	0.312**		0.075**	0.194***
(continuous treatment)		(0.030)	(0.063)		(0.093)	(0.144)		(0.033)	(0.070)
Share of 30-60km band in command area			-0.049			0.120			-0.182**
			(0.073)			(0.176)			(0.082)
Observations	26,856	72,732	72,732	26,856	72,732	72,732	24,618	66,671	66,671
R^2		0.70	0.67		0.82	0.80		0.15	0.06
Panel D. 40km radius									
	1	2	3	4	5	6	7	8	9
Command area in town catchment area	0.020			0.032			-0.002		
(binary treatment)	(0.019)			(0.034)			(0.026)		
Share of 0-40km band in command area		0.060**	0.068		0.442***	0.398***		0.048	0.134**
(continuous treatment)		(0.030)	(0.054)		(0.100)	(0.130)		(0.036)	(0.064)
Share of 40-80km band in command area			-0.012			0.068			-0.135*
			(0.066)			(0.146)			(0.075)
Observations	25,824	74,352	74,352	25,824	74,352	74,352	23,672	68,156	68,156
R^2		0.70	0.67		0.82	0.80		0.15	0.06

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

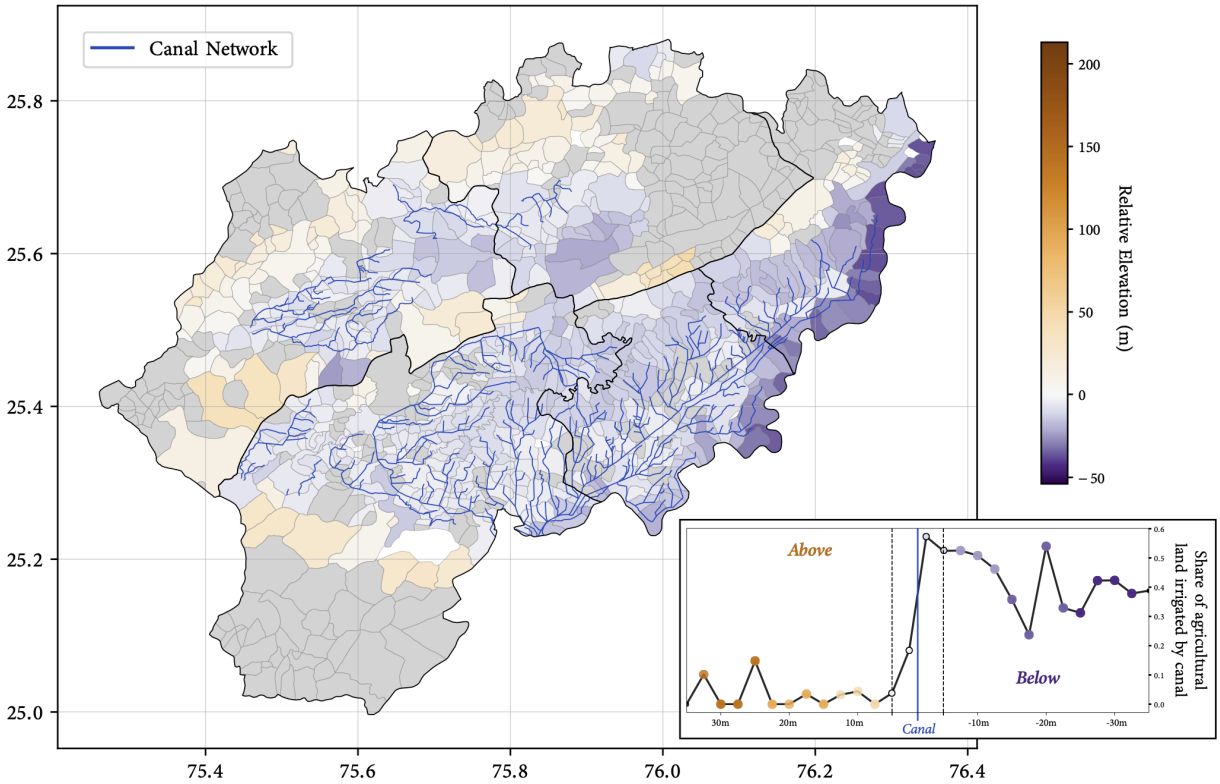
Notes: This table shows the effect of canal construction on town growth reported in Table 6 using various catchment area radii around the town results. An additional analysis includes the command area coverage of a second, outer band around each town as an additional independent variable. Panel B, using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

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 RD 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.