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

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

May 1, 2024

Abstract

We estimate the long-run effects of India's vast canal network, which provides irrigation water to 200+ million people. Canals reshaped Indian economic geography, with substantial economic changes both inside and outside of irrigated zones. Higher agricultural productivity raised population around irrigated villages, with no effect on village non-farm sectors. Structural transformation occurred almost exclusively through concentrated emergence and growth of towns. A model with mobile labor and urban non-farm productivity advantages rationalizes the findings. In the long run, the productivity effects of canals were equilibrated through the spatial reallocation of 50 million people, rather than through in situ structural transformation.

JEL Codes: O13, O18, Q15

*We are grateful for excellent research assistance from Sam Besse, Aditi Bhowmick, Ellie Baker, Kritarth Jha, Toby Lunt, Shraddha Mandi, and Sankalp Sharma. Gayatri Acharya, Richard Damania, Anju Gaur, Ijsbrand de Jong, Stuti Sharma, Vivek Srivastava and Esha Zaveri at the World Bank provided invaluable insight into the water and agriculture sectors in India, as well as assistance in obtaining data. We appreciate the feedback from seminar and workshop participants at UC Berkeley, Stanford, the University of Chicago, IIM Ahmedabad, IFPRI, BREAD, Dartmouth College, Warwick, Essex, Exeter, Geneva, the World Bank, STEG, NEUDC, Trinity College Dublin (TIME), UC Davis, UC Merced, Kent, IIES Stockholm, Birkbeck, Columbia, and George Mason. Financial assistance from the World Bank Agriculture Global Practice and Research Support Budget, Emergent Ventures (Mercatus Center, George Mason University), and the Weiss Fund, is gratefully acknowledged. Replication code can be found at `github.com/devdatalab/paper-india-irrigation-canals/`.

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

‡Tufts University, alison.campion@tufts.edu

§Tufts University, douglas.gollin@tufts.edu

¶Dartmouth College, paul.novosad@dartmouth.edu

1 Introduction

The link between agricultural productivity and structural transformation has long been a central concern of development economics (Nurkse, 1952; Lewis, 1954; Ranis and Fei, 1961; Schultz, 1964). Early authors such as Johnston and Mellor (1961) and Jorgenson (1961)—echoed later by Mellor (1986), Timmer (1988), and World Bank (2007)—argued that agricultural productivity growth was an essential precursor for broader structural transformation and long-run economic growth.[1] However, the literature offers both theoretical and empirical challenges to this view of the relationship between agricultural productivity and sectoral change.[2]

This paper studies one of the most significant episodes of agricultural productivity change of the past two centuries. India's massive irrigation canals—artificial channels that carry water into dryland areas for application to crops, primarily during the dry winter growing season—span over 300,000 km and serve over 130,000 villages, nearly one out of every four in India. Canals were historically the most important source of irrigation in India, and even in the 21st century they are second only to groundwater, providing water to agricultural areas with over 200 million inhabitants. In 2011, fully 57% of rural Indians lived within 10 km of a canal.[3]

These canals are a novel context for studying the long-run impacts of technical change in agriculture because (i) they cause changes in agricultural productivity with sharp spatial boundaries that are sustained for decades; and (ii) the majority of canals were built 30–100 years in the past. Other agricultural interventions tend to gradually diffuse across space, making it more difficult to study

[1]This early literature held that productivity growth in agriculture could have the seemingly paradoxical effect of shrinking the agricultural sector as a share of the total economy. Building on the insight that food is an essential good for the poor, agricultural development economists generated a class of models in which countries that are unproductive in agriculture must devote large shares of labor and other resources to meet their food needs. Schultz (1953) referred to this phenomenon as the "food problem". The same mechanism lies at the heart of more recent work, which relies on non-homothetic preferences as the main driver of structural transformation (Gollin et al., 2002, 2007; Alvarez-Cuadrado and Poschke, 2011; Comin et al., 2021). The link between agricultural productivity growth and structural change also emerges in other models where productivity growth leads to endogenous changes in the relative price of agricultural goods (Ngai and Pissarides, 2007).

[2]For example, Matsuyama (1992) showed that in an open economy, increases in agricultural productivity can cause *specialization* in agriculture via comparative advantage, while Bustos et al. (2016) show that the direction of local structural transformation can depend on the factor bias of the technical change in agriculture.

[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.

their effects over long time periods.

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

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

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

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

The RDD analysis tests for long-run differences between places that have direct access to canal irrigation and those that do not, but it does not account for spillovers. For example, spillovers could occur through linkages in labor and goods markets, leading untreated settlements — such as higher elevation villages and regional towns — to experience changes in demand for both agricultural and non-agricultural labor. Spillovers could also occur through hydrology: canals could recharge regional groundwater tables, increasing access to pump irrigation in the control settlements above the canal.

We employ two distinct empirical strategies to test for two different forms that spillovers may take. First, to measure spillover effects into untreated settlements near canals – the control group in the RDD – we compare settlements directly above canals to settlements that are in the same district but are more distant from canals. This setup only identifies spillovers that shrink with distance, but the likely spillovers — market linkages to canal-treated settlements and groundwater recharge enabling pump irrigation — do plausibly decay with distance due to spatial frictions.[4]

Second, to test for the possibility that economic growth arising from canals occurs through concentrated production clusters, we draw on a one hundred-year panel of population for 8,000 urban areas across India.[5] We use a difference-in-differences design that studies urban growth before and after regional canals are built, following De Chaisemartin and d'Haultfoeuille (2020).

The RDD analysis reveals sharply improved agricultural outcomes in the settlements directly treated by canals. Directly treated settlements have more land under cultivation, greater irrigated acreage, a higher likelihood of growing water-intensive crops, and higher estimated yields.[6] The yield effects are observed almost entirely in the relatively dry winter (*rabi*) season: canals improve water access in a second cropping season but generate much smaller differences during the summer monsoon (*kharif*) growing season, when rainfall is much more plentiful. There are no spillover effects on a range of agricultural outcomes: irrigation levels, yields, and land use in above-canal settlements

[4]We use entropy balancing (Hainmueller, 2012) to reweight the sample of distant and above-canal settlements, thus comparing settlements with similar distributions of natural characteristics (climate, topography, and agricultural potential).

[5]Urban population is the only high resolution variable that is available in the era of canal construction.

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

are highly similar to those in more distant settlements. The sharp differences in agricultural outcomes between above-canal and below-canal settlements have been sustained over many decades, making them a useful natural experiment for studying what happens to the rest of the economy in the long run following a major increase in agricultural productivity.[7]

The agricultural changes brought about by canals cause substantial population growth in irrigated regions, but ultimately little local structural change. Below-canal settlements have 22% higher population density; villages immediately above canals have 5% higher density. But below-canal, above-canal, and distant settlements have highly similar shares of workers employed in manufacturing, services, and even in agro-processing. There is evidently an increased demand for labor, as evidenced by higher population density in irrigated areas, but these highly agricultural settlements do not develop substantial non-farm sectors.

Structural change does take place, however, in the form of urban population growth. In the town panel, we find that after a canal is completed, towns are more likely to emerge in the region and existing towns grow more quickly in the following decades. These gains are concentrated in smaller towns, which are likely to have more of their economy directly tied to regional agriculture.

The net population movements are substantial in magnitude; a back-of-the-envelope calculation suggests that India's canal network has increased the population of canal-proximate settlements by about 48 million people and added 5 million people to canal-region towns. To put these numbers into perspective, the Partition of India and Pakistan resulted in the displacement of 17 million people, while the largest episode of international migration in history — from Italy to the New World in 1880–1915 — involved approximately 13 million people. While our high-resolution data do not directly record population flows across space, multiple pieces of evidence suggest that migration accounts for a large share of the observed population changes. In short, India's canal network has had a massive effect on its economic geography.

Canals have heterogeneous effects on living standards. Using small area estimates from household asset and earnings data (Elbers et al., 2003), we find that canals produce no persistent consumption

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

gains for the roughly 70% of rural households who own little to no land. In contrast, households in the higher quartiles of landholding show substantial increases in consumption in directly-irrigated settlements, with effects increasing in the size of land holdings. There are no consumption spillovers into above-canal settlements, suggesting that these gains are driven by higher returns to land.

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

This paper extends a substantial literature linking technical change in agriculture to structural change. Some studies (Foster and Rosenzweig, 1996, 2004b; Hornbeck and Keskin, 2015) have failed to find local impacts of agricultural productivity growth on the non-agricultural sector. Other analyses (Bustos et al., 2016; Gollin et al., 2021) have found structural transformation effects at the regional and national scales. We show that limited local structural change can be consistent with significant transformation at a larger geographic scale. In our context, regional urbanization is essential for understanding the effects of agricultural productivity change, recalling Bustos et al. (2020), who found that land rents from technical change in agriculture were invested in cities.[8]

[8] An example of this capital channel is discussed at length in the context of colonial Bengal in Bose et al. (1993). Also consistent with our results, Liu et al. (2023) find that long-run agricultural productivity losses from higher temperatures due to climate change dampened non-agricultural employment and urbanization in Indian districts. Our results also echo Foster and Rosenzweig (2004a), who argued that agricultural productivity shocks have substantially different effects on landowners and the landless. There is a large body of evidence on responses to transient agricultural productivity shocks due to weather (Adhvaryu et al., 2013; Colmer, 2021). Emerick (2018) and Santangelo (2019) in particular find that non-tradable employment increases in districts experiencing positive agricultural productivity shocks, consistent with our model of demand-driven structural change. Our paper speaks less to this literature

Our findings are notably different from a concurrent partial equilibrium analysis of canals by Blakeslee et al. (2023), who use an RDD identification strategy similar to the first part of our analysis. Like us, they find population growth and little structural change in villages, but they do not study effects beyond directly-irrigated canal command areas. They argue that the population increases in canal villages came at the expense of command-area towns, which they estimate lost over 20% of their population. They conclude that canals ultimately caused a net decrease in regional population. This result comes from comparing urban populations inside and just outside of canal irrigated zones. But towns do not grow crops and thus do not benefit from the direct effects of irrigation; instead, as our model shows, they have market linkages to agricultural regions and thus urban growth can appear anywhere in the surrounding region — not necessarily in irrigated zones. Our analysis shows that in fact, canals generated a very large net population increase — over 50 million people — into both canal-irrigated rural areas and nearby (but largely unirrigated) urban areas. Only half of these rural changes (and none of the urban changes) are detectable at canal boundaries, exactly as our spatial equilibrium model predicts. Our empirical approach thus highlights the importance of taking spatial equilibrium seriously and studying the broader economic geography of local shocks.[9]

We also contribute to the literature on how labor flows respond to economic shocks in both high- and low-income countries (Greenstone et al., 2010; Allcott and Keniston, 2018; Imbert and Papp, 2020). Recent empirical work focused on causal identification has often studied competition for workers between the farm and non-farm sectors in models and short-run contexts where the labor mobility channel plays only a small role.[10] Our analysis suggests that migration may be the primary long-run adjustment channel to agricultural change. Indeed, the very nature of structural transformation around the world

on transient shocks because we study how people adjust to large, permanent changes in agricultural productivity.

[9] Specifically, Blakeslee et al. (2023) conclude that canals result in urban depopulation, because towns in canal-irrigated areas are smaller in equilibrium than towns just outside of irrigated areas. This result did not replicate in our data, and replication files or data from the other paper are not available. However, we note that their result could also be explained by the emergence of new small towns close to canals (a result that we find), leading to a decrease in average town size — an example of Simpson's paradox. Our analysis of towns is fundamentally different from Blakeslee et al. (2023), in that we test whether towns have emerged and grown in the broader regions of canals. We do not estimate the elevation RDD for strictly urban locations because: (i) it would be biased by town emergence; (ii) the town sample is too small for RDD estimation with any precision; and (iii) theory suggests the primary impact of canals on urban spaces (i.e. through regional market linkages) does not depend on whether the urban land itself can be irrigated with canal water.

[10] 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.

has involved the movement of billions of people from farms to cities, sometimes across large distances.[11]

Our results further highlight the high barriers to rural industrialization. Asher and Novosad (2020) and Burlig and Preonas (2022) find that major investments in rural roads and electrification respectively have generated limited effects on non-farm activity in India.[12] Faber (2014) finds that highway construction through peripheral areas in China in fact caused deindustrialization. These papers suggest that while infrastructure investments in rural areas may improve well-being, they often do *not* cause substantial changes in *in situ* non-farm opportunities.

Our work also adds to a growing literature estimating the impacts of access to irrigation (Duflo and Pande, 2007; Sekhri, 2014; Blakeslee et al., 2021; Jones et al., 2022). Finally, our paper contributes to recent work highlighting the fact that spillovers from treated to untreated units can be important components of overall treatment effects (Miguel and Kremer, 2004; Egger et al., 2022; Adao et al., 2022). Indeed, over 40% of the net population flows induced by canals occurred outside of directly-irrigated villages.

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

2 Context

As a semi-arid region with a highly variable monsoon climate, South Asia has long depended on irrigation for its agricultural productivity. For much of history, this has primarily involved gravity flow surface irrigation through canals of various types. At the end of the 19th century, India had 12 million hectares of irrigated land — four times more than the United States and six times more than Egypt (Shah, 2011). The British oversaw the construction of vast canal networks, often privately funded and yielding high returns, until the end of the Raj in 1947. Canals were used to divert water from India's

[11] While there is a widespread idea in the literature that permanent migration in India is rare, this idea is focused on the set of rural men who migrate for work. Over 25% of women have changed location of residence at least once in their lives, and lifetime migration rates for men approach 15% (Kone et al., 2018).

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

major rivers to its arid regions, where they facilitated settlement of otherwise uninhabitable land. The best known example was the construction of canals into low-rainfall regions of western Punjab (now in Pakistan), creating nine distinct "canal colonies" in regions that had not previously supported much settled cultivation. The canal colonies covered some 2.5 million ha of land and eventually absorbed about one million new migrants into both rural and urban areas. The city of Lyallpur (now Faisalabad) was a direct and intended product of the canal construction (Douie, 1914). In most cases, however, irrigation canals did not open land on the "far" extensive margin. Instead, the goal was to improve the agricultural potential for communities already engaged in settled agriculture (Stone, 1984).

After gaining independence, the Government of India prioritized canal-building, seeking to avert mass hunger during a period of high population growth (Mukherji, 2016). Later, canals were built to provide irrigation for the input-intensive high-yielding varieties of food crops that powered India's Green Revolution.

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

Key to our empirical strategy is the fact that canals are costly investments whose exact routes are difficult to modify for political or other considerations. Scholarship on colonial canal construction emphasizes how tight budgets and topography dictated the local feasibility and routing of canals (Stone, 1984). Similar constraints shaped post-independence canal placement; for the Kosi canal in Bihar, built from 1953–1964, officials lamented that 246,000 acres of land with high irrigation potential could not be irrigated without extensive leveling, for which there was insufficient money and machinery (Pant, 1981). Indeed, Shah et al. (2001) argues that the rise of groundwater irrigation and relative decline of canals was due in part to the inability to target canal placement: "compared to large surface systems whose design is driven by topography and hydraulics, groundwater development is often much more amenable to poverty-targeting."

Figure 1 shows the distribution of official completion dates of the major and medium canals studied here.[13] A caveat to this figure is that the official "completion date" is updated if a canal undergoes a substantial renovation, such as projects funded by the Accelerated Irrigation Benefit Program. As a result, the older dates (when India had fewer canals) mostly represent original canal construction, whereas many of the recent dates in fact reflect rehabilitation projects on canals built several decades earlier. Construction rates increased following India's independence in 1947, although post-independence canals were generally shorter than those constructed under the British Raj in the 19th and early 20th century. By 2011, 51% of India's 600,000 settlements were within 10 km of a major or medium irrigation canal, with a median canal construction start year of 1972 and completion year of 1980. Given that our primary outcomes are measured in 2011–2013, the canals in our study are typically at least thirty years old.

It is worth noting that the effects of canals that we document in this paper were understood and documented by contemporary observers. Colonial officials understood that canals drew labor from other regions to work in newly-irrigated fields. Canal irrigation increased the returns to weeding, hoeing, and intensive crop management. In one of the few works of economic history dealing directly with canal irrigation, Stone (1984) offers numerous examples from contemporary colonial records to show that canal irrigation altered patterns of labor use in rural villages, inducing shifts in cropping patterns towards high-value water-intensive crops (in that era: sugarcane, cotton, indigo, and wheat) and away from dryland crops such as sorghum and millet. A 19th century colonial document notes that "hire rates [i.e., wages] in canal villages tended to be slightly above those prevailing in well villages [...] a canal village presents a richer appearance than a well village" (Stone, 1984). Colonial documents similarly note that higher wages and year-round labor demand frequently induced a sometimes-sticky inflow of seasonal workers from rain-fed regions, especially in times of drought.[14]

[13]Major canals are defined as serving 10,000 or more hectares, while medium canals serve areas of 2,000–10,000 hectares. Canals serving fewer than 2,000 hectares are termed minor canals and are not included in the database of canals at the heart of this study.

[14]One colonial official, writing in 1873 about Muzaffarnagar District, observed, "The statistics of the tract when

These historical features of irrigation canals motivate our theoretical framework and empirical strategy.

3 Model

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

However, the relationship between agricultural productivity gains and structural transformation has been shown to depend on assumptions relating to the openness to trade (Matsuyama, 1992), the substitutability of agricultural and non-agricultural goods (Ngai and Pissarides, 2007), the factor intensity of technological change (Bustos et al., 2016), and capital mobility (Foster and Rosenzweig, 2007; Bustos et al., 2020), among others. We write a parsimonious model that deviates from existing models in the literature in two key dimensions that reflect our empirical context. First, we model an economy in which labor flows freely across space in the long run, but not in the short run. This offers a contrast to many models that allow for labor mobility across sectors but not across locations.[15] Second, we allow for spatial variation in non-agricultural productivity, such that larger settlements have a productivity advantage in the production of non-agricultural goods.

Our model is based on several empirical features of India's agricultural system. India's rural economy is reasonably characterized as a large number of predominantly local sub-economies that

examined in detail show clearly enough that... population has increased in a marked manner in this tract only in those estates which are sufficiently watered by the canal" (Cadell, 1873). Stone (1984), reviewing a wide set of original source material, notes a distinct "shift of population to canal villages."

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

are embedded in a larger national economy. Each rural region features an expanse of agricultural land, divided into villages, typically with a larger market town that serves as an economic center.[16] Agricultural land is most often privately owned and managed. Most farms are small (Foster and Rosenzweig, 2017), and farmers may hire labor from a large pool of landless workers. These observed features of the data give shape to our simple model.

3.1 Model Setup

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

The economy produces two goods: an agricultural good a that is traded beyond the region and a non-agricultural good c that is costlessly traded within the region but non-tradable beyond the region. The non-tradable good might correspond to services, such as haircuts; but it could also represent manufactured goods with low value per unit transport costs, such as bricks.[17]

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

We consider three periods. In the initial period, the region is in a long-run spatial equilibrium with the rest of the country. Following the initial period, a canal is built that raises agricultural

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

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

productivity in village 1 but not village 2. During the second period, which describes the short run, labor is mobile across sectors and settlements within the region, but not between the region and the rest of the country. In the third period, which we call the long run, labor is also mobile across regions, generating a new spatial equilibrium.

3.1.1 Preferences and utility

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

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

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

3.1.2 Production and trade

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

Each of the region's villages has an endowment of land (L_i) and labor (N_i), while the town has only labor (N_0). The regional economy has a labor force of N people, where N_i is the labor force of village v_i and N_0 is the labor force of the town. Thus, $\sum_i N_i = N$. The supply of land is fixed in all periods, while the total regional labor force N is fixed only in the short run following canal construction. For simplicity, we assume that all land in the region is held by a single landowner, who resides in the town and receives all land rents. All individuals supply one unit of labor to the market, inelastically.[18]

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

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

to natural advantage or agglomeration economies, the town has the highest C_i in the region. Recall that the traded good m is consumed but not produced within the region.

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

3.2 Equilibrium

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

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

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

We begin by solving for the equilibrium with full labor mobility, which characterizes both the initial period and the long-run equilibrium following canal construction.

3.2.1 Equilibrium with full labor mobility

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

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

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

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

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

The total agricultural output is thus:

$$Y_a = Y_{a1} + Y_{a2} = \left(\frac{\theta}{w_{LR}}\right)^{\frac{\theta}{1-\theta}} \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2\right]. \tag{3.4}$$

Combining equations 3.2 with 3.4, we can solve for the non-tradable labor force:

$$N_0 = w_{LR}^{\frac{1}{\theta-1}} \theta^{\frac{\theta}{1-\theta}} \left(\frac{\beta}{1-\beta}\right) \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2\right]. \tag{3.5}$$

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

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

3.2.2 Equilibrium with only local labor mobility

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

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

$$w = N^{0^{\theta-1}} \theta \left[\left(\frac{1}{\theta} \right) \left(\frac{\beta}{1-\beta} \right) + 1 \right]^{1-\theta} \left[A_1^{\frac{1}{1-\theta}} L_1 + A_2^{\frac{1}{1-\theta}} L_2 \right]^{1-\theta} \tag{3.6}$$

Comparative statics in the short run: As above, we assume canal construction raises A_1. The change in the wage is given by

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

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

3.2.3 Summary

The model illustrates the two major contributions of this paper. First, the long-run impacts of agricultural productivity gains are geographically heterogeneous: directly-treated villages gain agricultural workers while non-agricultural growth occurs in places that have a comparative advantage in that sector — urban areas, in our model. Second, the long-run impacts of agricultural productivity shocks are different from the short-run impacts due to the mobility of labor. Because the irrigation canals that we study were built so long before the collection of available high-resolution data needed to study their effects, our empirical analysis focuses on the long-run equilibrium after canal construction.

4 Data

We assemble recent high-resolution data on the universe of firms, households, and locations in India, building on data from the SHRUG open data platform (Asher et al., 2021). Because the reclassification of rural villages into urban towns is an endogenous outcome driven by population density and administrative discretion, we define a "settlement" (a municipality, either a village or town) as our primary unit of observation. The analysis dataset covers 590,000 settlements (8,000 are towns; the rest are villages), which are nested in 5,000 subdistricts and 700 districts. Most outcome data is from the period 2011-2013, while the panel of urban population covers the period 1911-2011.

The 2011 Population Census provides demographic variables for every settlement and data on cultivated and irrigated land area in every settlement in India. The census also records the three main crops grown in each village, from which we create an indicator for villages that grow a water-intensive crop (cotton, sugarcane, or rice).[19] Since settlements are heterogeneous in size, our preferred measure of population is density, which we define as inhabitants per square km.[20] While these data are cross-sectional for all settlements in 2011, we also have a decadal panel of town population (i.e. excluding villages) extending back to 1911.

The 2012 Socioeconomic and Caste Census (SECC) is an asset census that was undertaken in all Indian villages to determine eligibility for means-tested programs. From SECC microdata, we generate the share of adults aged 20–65 who have completed primary, middle, and secondary school, as well as predicted consumption per capita using small area estimation based on the income and asset variables in the SECC.[21] Because the SECC is recorded at the household level, we can calculate these outcomes separately for landowners and landless households. We collapse all SECC measures to the settlement level for the analysis.

The 2013 Economic Census is a complete enumeration of all non-farm economic establishments in India, which we use to measure non-agricultural economic activity for each settlement. We calculate employment as a share of the 2011 Population Census adult population.[22] We use the National Industrial Classification codes of firms in the Economic Census to calculate the share of the adult

[19]As Population Census data on agricultural outcomes are available only in villages, analysis of these outcomes excludes towns.

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

[21]The latter follows the methodology of Elbers et al. (2003) and is described in detail in Asher and Novosad (2020). 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.

population employed in manufacturing, services, and agro-processing in each settlement.[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 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 both for the monsoon (*kharif*) season, which runs from late May through early October, and for the winter (*rabi*) season, late December through late March (Selvaraju, 2003). For each season, we define productivity by subtracting the mean of the first six weeks of the season from the maximum EVI value reached over 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 include 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]

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] The command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal, extending out from the canal and ending at a boundary that is determined by a combination of canal flow, terrain, and soil type. These data allow us to calculate settlement-level measures such as distance to the nearest canal, elevation relative to the nearest canal, distance to the nearest river and coastline, and whether or not the settlement is located inside a command area.[26] The WRIS also provides dates of canal construction and completion; however, our research on individual canals suggests that recent start and

[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. Agro-processing is defined as a subset of manufacturing employment codes, specifically NIC codes 10 and 12.

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

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

[26]For distance measures, we characterize the each settlement's location with its centroid.

end dates in WRIS often represent canal rehabilitation efforts, rather than new canal construction.[27] It is therefore challenging to identify exact construction dates of what appear to be more recent canals. Older construction dates appear to be more credible, as canal investments in the post-independence period and earlier were more often new canals rather than maintenance of existing infrastructure. As a result, we are unable to identify canals that were actually built in the two decades preceding our outcome data, preventing us from estimating the short-run effects of the arrival of canal irrigation.

Using settlement polygon GIS data from ML Infomap, we extract the distribution of elevation in each settlement from Shuttle Radar Topography Mission (SRTM) raster data. Following Riley et al. (1999) and Nunn and Puga (2012), we calculate the ruggedness of a settlement'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 the same settlement polygons, we generate various settlement-level geophysical characteristics that could correlate with canal placement and agricultural productivity. We extract 10-day rainfall values from the Climate Hazards Center InfraRed Precipitation with Station (CHIRPS) dataset (Funk et al., 2014) to generate mean, total annual rainfall from 2010–2014 as our rainfall measure. Similarly, we extract monthly maximum daily temperature for each settlement from the Climate Hazards Center Infrared Temperature with Stations (CHIRTS) dataset (Funk et al., 2019), then compute the average maximum monthly temperature over the 2010–2014 time period as our temperature measure. We measure agricultural productivity potential for India's two largest crops (rice and wheat) from the FAO Global Agro-Ecological Zones (GAEZ) database. Finally, our soil quality measure is published by the Harmonized World Soil Database and describes the rooting condition of the soil (Fischer et al., 2008). Rooting conditions reflect the soil depth, volume, and presence of gravel that can all impact the ability of crops to effectively gain a foothold, take-up nutrients, and grow to their peak yield potential. We define the binary variable as 1 indicating slight or no limitations of rooting

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

conditions (80–100% of potential quality) and 0 indicating moderate to severe limitations.

To test for balance, we would ideally like to compare treatment and control settlements prior to the construction of the first irrigation canals. The 1951 Population Census is the earliest data source that we have for making detailed comparisons of this kind, and it predates many of the canal projects in our data. We were able to digitize village tables from this census from archived District Handbook PDFs for six states (Gujarat, Karnataka, Madhya Pradesh, Maharashtra, Rajasthan, and Uttar Pradesh). We extracted and matched data from 31,533 villages, 4172 of which match our analysis sample and are proximate to canals that were built after 1951.[28] We were able to construct the following variables: total population, male-to-female sex ratio, population density, mean household size, and literacy rate.

Finally, to test for the effects of canals on migration, we use data from the 1987–88 (43rd) round of India's National Sample Survey, which collected data on a sample of households towards the end of India's major post-independence era of canal construction. We define in-migrants as respondents who reported having had a previous place of residence different from their current one.

For a detailed description of every variable used in this paper, please see Appendix Table A1.

5 Empirical Strategy

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

[28]The match rate is relatively low due to the low scan quality of many of the district handbooks, combined with the challenge of matching villages by name to the 1991 Census (the nearest year which links to our sample.

growth, we use a hundred-year panel of town populations and a difference-in-differences estimator.

5.1 Regression Discontinuity Estimates of the Direct Effects of Canals

Canals provide water to fields through a system of gravity-driven secondary canals, trenches, and pipes. Because water delivery depends physically on gravity, fields must be at a lower elevation than a canal in order to be irrigated with canal water; settlements above the canal will not benefit directly. Our main identification strategy compares settlements close to canals with elevations that put them either just above or just below the threshold that would give them access to canal water. For this analysis, below-canal settlements are considered treated by canals and above-canal settlements serve as controls. As discussed in Section 2, canals are difficult to target locally and thus our treatment and control settlements are likely to meaningfully differ only in that treatment settlements receive large amounts of canal water and control settlements do not.

A settlement polygon is characterized in the data by a distribution of elevation values from the set of pixels within its borders. We define the polygon elevation as the 5$^{\text{th}}$ percentile of the polygon's pixel distribution; this value strongly predicts the difference in canal irrigation between treatment and control areas (see Appendix Figure A1).[29] For each settlement, we also calculate the elevation of the canal at its nearest point to the settlement.

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

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

where $y_{i,s}$ is an outcome in settlement i and subdistrict s and $REL\_ELEV_{i,s}$ is settlement elevation minus canal elevation (such that a negative value means that the settlement lies below the canal, and thus can receive its water), and $X_{i,s}$ is a vector of geophysical controls (ruggedness, mean annual rainfall, maximum annual temperature, distance to the nearest river, distance to the coast, the GAEZ crop

[29]Results are similar if we use the 25$^{\text{th}}$ percentile or median elevation to define above/below canal thresholds (Appendix Tables A5, A6, A7, and A8). We chose the 5$^{\text{th}}$ percentile in order to have a control group with close to zero canal irrigation; when we estimate spillover effects below, interpretation is most straightforward if the above-canal group experiences no direct treatment by canal water.

suitability measure for irrigated rice and wheat, and a soil quality measure of rooting conditions).[30] 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 total population averaging approximately 250,000 people. Standard errors are clustered at the subdistrict level to account for spatial correlation.[31] In the absence of spillovers to untreated settlements, the effect of canal irrigation is captured by β_1, which is the difference in outcomes between settlements just below and just above the canal. Appendix Figure A2 shows a map of a single district, along with its canal network, elevation profile, and an analog of the first stage RDD graph showing the share of land irrigated by canal.

The main analysis sample includes settlements within 10km of distance and 50m of vertical elevation from the nearest canal.[32] As our outcome data is from 2011 onwards, we exclude from our analysis sample any settlements whose closest canal is listed as incomplete as of 2011. We limit the sample to subdistricts that have at least one settlement in both the treatment and control group. Settlements with elevation very close to the treatment threshold have an ambiguous treatment status — for example, a settlement could have some of its land above the canal (and thus not treatable with canal water) and some of its land below the canal (and thus treatable). Inclusion of these settlements would bias RDD estimates toward zero; we therefore exclude a "donut hole" of settlements within 2.5m in elevation of the nearest canal in either direction. Finally, to avoid comparing lowland irrigated areas with rugged hilly areas, we impose a balance restriction on the Terrain Ruggedness Index (TRI). We allow a maximum 25% difference in mean TRI between below-canal and above-canal settlements in a given subdistrict; if the percent difference is greater, the entire subdistrict is dropped from the sample. Table 1 shows the sample size and mean values for all variables used in our analysis after each stage of the sample selection. We use the ruggedness-balanced analysis sample (Column 4) for our primary analysis, but show robustness in the Appendix to alternate sample definitions (Appendix Tables A5–A8). The ruggedness-balanced

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

[31]We show robustness to the use of Conley (1999) standard errors in Appendix Tables A5–A8.

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

analysis sample is representative of the universe of settlements in India on most dimensions: around half of agricultural land is irrigated, about 60% of village land is dedicated to agriculture, there is approximately 1 non-farm job for every 10 adults, and just under half of adults have completed primary school.

RDD validity requires that there are no pre-treatment differences at the threshold between above- and below-canal settlements. Since canal infrastructure in India was built throughout the 19th and 20th centuries, and treatment status is determined at the settlement level, there are no comprehensive high-resolution socioeconomic or agricultural data available to test this assumption. However, we can test for differences in time-invariant geophysical measures, which could proxy for natural advantages that might have affected canal placement and economic outcomes. Table 2 shows estimates of Equation 5.1 on geophysical fundamentals (with the specific outcome excluded from $X_{i,s}$ in each regression), demonstrating that there are no significant differences between above- and below-canal settlements in ruggedness, distance to coast, soil quality, average annual rainfall, or crop suitability for rice or wheat. We do find small imbalances on temperature and distance to rivers. The temperature difference is tiny in magnitude (0.037 degrees Celsius, on a mean of 32.54, a 0.1% difference) and would if anything lower agricultural productivity in below-canal settlements, as higher temperatures in India reduce agricultural yields (Colmer, 2021). Canal villages are also somewhat (1.5 km, or 6%) further from rivers. We control for all of these geophysical variables in all of the regressions below.

We also conduct balance tests using village-level demographic data from the 1951 Population Census. To do this we scraped and digitized the District Handbook PDF files for 32,765 villages in 109 districts across six states (Gujarat, Karnataka, Madhya Pradesh, Maharashtra, Rajasthan, and Uttar Pradesh). We present the results in Appendix Table A2. We find no evidence of imbalance across any of the five variables that we are able to construct (log population, sex ratio, population density, household size, and literacy rate), although we acknowledge that our limited sample may restrict our ability to identify differences.[33]

As a robustness exercise, we use a secondary regression discontinuity design that compares

[33] While it would be desirable to test for balance in 1991 (where we have settlement-level data) for canals built after 1991, this is not possible for two reasons. First, as discussed in Section 4, the vast majority of canals were built before 1991. Second, because the WRIS data does not distinguish canal rehabilitation dates from completion dates, many canals with post-1991 dates were in fact built much earlier.

settlements just inside and just outside of the canal command area.[34] We define the running variable as the distance between settlement centroid and command area boundary, defining it negatively inside the command area.[35] The estimation is otherwise similar to that above, but we additionally divide each command area boundary into 10km segments and include a fixed effect for each segment, ensuring that we are comparing settlements across the same stretch of each command area. Standard errors are clustered by these segments. This strategy exploits the variation in the xy-plane, whereas the primary (relative elevation) strategy exploits variation in the z-axis. The identifying assumption is that settlements just inside and just outside the command area boundary would have similar outcomes if the canal had not been built. While the command area definition may exploit finer details of local topography, we prefer the relative elevation strategy, as boundaries of command areas may be subject to some discretion by officials, who might have incentives to finish canal branches in particular places or to mark one settlement or another as within the official command area.[36] We test for balance with this command area boundary strategy in Appendix Table A4, finding no evidence for any imbalance, apart from a very small difference in temperature.

5.2 Testing for spillovers into above-canal areas

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

[34] This is similar in design to the strategy used in concurrent work by Blakeslee et al. (2023). Recall that the command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal.

[35] 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.

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

We test for local spillovers by testing for differences in economic outcomes between RDD control settlements and an alternative sample of control locations: distant settlements within each district, which lie 15–50km from the nearest canal but have similar geophysical characteristics to the above-canal settlements that serve as the control group for the RDD. This strategy is predicated on the assumption that any mechanism driving spillovers is likely to decay with distance from treated areas. If spillovers do not decay over distance, they are more difficult to measure. For example, if landless labor were perfectly mobile across all of India, then a new canal could have a small positive impact on wages in the entire country, but there would be no control group against which such an effect could be measured. While we cannot rule out universal effects like these, our empirical design will identify the existence of spillovers as long as they have a non-zero gradient in distance. Given the nature of the likely spillovers (groundwater recharge and market linkages from canal-treated areas) and India's high spatial frictions from factors like poor transportation infrastructure, language barriers, and barriers to trade across states, we consider it improbable that spillovers will extend equally across the entire country.

To ensure that the distant settlements are a reasonable counterfactual to canal-proximate settlements, we use entropy balancing (Hainmueller, 2012) to assign weights to settlements to minimize differences in distributions (first, second, and third moments) of geophysical fundamentals in distant, above-canal, and below-canal settlements.[37] Entropy balancing is a useful and increasingly popular matching method because it does not impose functional form assumptions on propensity weights and thus achieves better balance than propensity-score matching.[38] Following the literature, we enforce common support by dropping outliers (the top and bottom 2.5% for each of the matching variables). We test for spillovers using the following estimating equation:

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

where below-canal settlements are defined as in Section 5.1 and the distant settlement group is all

[37] The geophysical variables used in the entropy balancing estimator are the same set that are used as controls in all regressions: ruggedness, rainfall, maximum annual temperature, distance to the nearest river, distance to the coast, crop suitability for irrigated rice and wheat, and soil quality.

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

settlements 15–50km from a canal.[39] Above-canal settlements 0–10km from the nearest canal are the reference group. The coefficient γ_2 describes the difference between the omitted group (the above-canal settlements) and the distant settlements. If there are meaningful spillovers from canal-irrigated areas into proximate untreated (above-canal) settlements, we expect γ_2 to be significantly different from zero. $X_{i,d}$ is the same vector of time-invariant geophysical controls as in the RDD specification above. The sample of above- and below-canal settlements is the same as in the RDD. To compare that there are sufficient distant villages in the sample, 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. In tables, we report $-\gamma_2$, such that it describes the effect of being in a canal spillover zone.

Note the difference between γ_1 here and the RDD estimate of β_1 from Equation 5.1. The RDD estimate describes the local difference *at the elevation threshold* between above- and below-canal settlements; γ_1 is the estimate of the average difference between below-canal settlements and above-canal settlements. If there is no relationship between the RDD running variable (elevation) and the outcome, then we will find $\gamma_1 = \beta_1$. In practice, the RDD estimator β_1 requires weaker assumptions for causal interpretation than γ_1 and is thus a better estimator of the direct effects of canal irrigation.

5.3 Town growth over time

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

To test whether canals affect regional urbanization, we instead exploit variation in canal construction dates and examine whether town population growth changes following the construction of nearby

[39] In robustness tests, we vary the distance criteria of the distant settlements.

canals.[40] The available data (from the 2011 Population Census) records the population of each 2011 town in each census year going back to 1901, beginning with the first year in which the Census defined a location to be urban.[41] Such an analysis is not possible for any other outcome, because urban population is the only variable available in a long panel that spans the many decades of canal construction.

To define whether a town is near a canal, we first draw a circle with a 20 km radius around each town.[42] We define a continuous measure of canal treatment ($CANAL\_SHARE_{i,t}$) for town i in year t as the percentage of the circle area that is overlapped by canal command areas. An alternate specification defines a binary treatment variable that takes the value 1 if more than 20% of the circle is covered by canal command areas.[43]

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:

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

Outcome $y_{i,t}$ is either an indicator for town existence or log(town population) in town i in year t, and ζ_i and ν_t are town and year fixed effects, respectively. When $y_{i,t}$ represents population, we assign the population value 2000 to towns that do not yet exist — this treats settlements before they become towns as if their size was just below the average population at which towns first appear in the data.[44] For the binary treatment, we use the estimator from De Chaisemartin and d'Haultfoeuille (2020), using the not-yet-treated towns as the control group and defining the treatment year as the first year when a town's 20 km radius catchment area is more than 20% covered by canal command

[40] Population growth is widely used in the economic history and urbanization literatures to proxy for overall economic growth (Ashraf and Galor, 2011; Hanlon and Heblich, 2022).

[41] Locations in India are considered urban when they meet the following three criteria: a) the population exceeds 5000, b) more than 75% of the male workforce is employed in the non-agricultural sector, and c) the population density is over 400 per square km. We do not observe former towns which do not exist any longer, but given India's rising urbanization, town disappearance is very rare.

[42] On average, there are 90 villages within 20 km of each town.

[43] In the era of canal construction, a straight-line distance of 20 km represented a multi-hour journey in much of India. We show that results are robust to different radius lengths and treatment thresholds.

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

areas. Standard errors are clustered at the district level.

6 Results

6.1 Direct Treatment Effects of Canals: Regression Discontinuity Estimates

We first report RDD estimates of the direct effects of canal access on irrigation outcomes, the mechanism through which we expect all other equilibrium effects to occur. Panel A of Table 3 shows that in canal-treated areas, 7.5 percentage points more of the land under cultivation is irrigated (17.5% more than in control settlements), and 9.9 percentage points (309%) more land is irrigated by canals. There are no discernible changes in other sources of irrigation. We test separately for effects on tubewell use, which would suggest greater groundwater access (for example, if canals recharge aquifers) and find no effects in the RDD.

Panel B in Table 3 reports direct effects of canal access on agricultural outcomes. As expected, canal-treated settlements experience higher agricultural productivity, with much larger and highly significant effects in the relatively dry winter (*rabi*) growing season (7.1%, $p < 0.001$) than in the rainy (*kharif*) season (1.7%, $p = 0.062$). Settlements below canals also cultivate 2.7 percentage points more of their total land area, a 4.5% increase over control settlements, and are also 5% more likely to list a water-intensive crop (rice, cotton, or sugarcane) as one of their three primary crops. We find no evidence of increased capital intensity in agriculture, as measured by the share of households owning mechanized farm equipment.

The key question of this paper is how these major changes in agricultural productivity affect living standards and the growth of the non-farm economy. Panel C presents estimates of the impacts of canals on population density, non-farm employment, and predicted consumption. The only significant effect is on population: by 2011, treatment settlements have 15.4% more people per square kilometer than control settlements. Despite large gains in the productivity of the dominant economic sector in villages and to population, we find no significant difference in living standards between above- and below-canal villages. The point estimate on log consumption is +0.007, with a 95% confidence interval of $[-0.004, 0.018]$: we can rule out even small effects. There is also no evidence of structural transformation as measured by non-farm jobs per adult; nor do we find significant effects when we isolate manufacturing or even agro-processing, the sector with the most direct linkage to agricultural production (Appendix Table A3). Total non-farm employment is higher than in canal settlements (as would be expected given the increase in population) but the non-farm *share* of the economy (the outcome of interest) is unchanged. We do find a marginally significant positive effect on the service sector share. It is economically very small — about 5% of the control group mean. Canal settlements have higher human capital (Panel D of Table 3); we measure a small but precise increase in the share of the adult population that has completed primary, middle, and secondary school, as well as the population literacy rate.

Figure 2 shows regression discontinuity binscatters of key outcomes in each of the categories above, with outcomes residualized on fixed effects and geophysical controls, showing the treatment effect at the RDD threshold, providing clear visual evidence of the effects of canals on agricultural outcomes and population density but also no discernible jumps at the running variable threshold in employment and 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. The effect on population density is substantively larger than any other non-agricultural outcome.

The model in Section 3 suggests that the long-run spatial equilibrium will be characterized by equalization of returns to mobile factors (such as labor), but not to fixed factors (such as land). In the absence of high-resolution data on wages and land rents, we proxy the returns to these factors by estimating canal treatment effects on predicted consumption separately for landless households (who own only labor) and for land-owning households (who own both land and labor).[45]

The results on land ownership are presented in Figure 4 and Table 4. Panel A of Table 4 shows a 2.7 percentage point decline in the share of the population that are landowners in canal settlements relative to control settlements, with the average landholding size of landowners unchanged. This implies that the population increase in below-canal settlements is disproportionately driven by an increase in the number of landless households. The consumption effects of canals are substantially different for landed and landless households (Panel B of Table 4): there are no significant consumption effects

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

for landless households, but landowner consumption is 2.1% higher in below-canal settlements; this result is statistically significantly different from the estimate for landless consumption at the 1% level.

Partitioning landowners by nationally-defined landholding quartiles, effects increase monotonically by quartile, with no significant consumption effects on those owning <1 hectare of land (the 1st quartile), and a 2.9% effect on consumption for those in the top quartile owning >4 hectares (Panel B).[46] Both landless and landowning households experience gains in educational attainment, but effects for landowners are two to three times higher than for the landless (Table 4 Panel C). In short, the results are consistent with a model where the agricultural productivity gains from canals draw in new landless labor until a spatial equilibrium is reached, with equal returns to labor in areas above and below the canal, as we discuss further in Section 7.

6.1.1 Robustness

The RDD results are robust to alternate parameter choices. To show robustness, we replicate all of our primary outcomes in Appendix Tables A5–A8; the different panels of the table show the result of different specifications for each outcome. Panel A shows results when we remove the ruggedness balance restriction, and include imbalanced subdistricts. Panels B shows results where settlement elevation is defined as the 25$^{\text{th}}$ percentile pixel, rather than the 5$^{\text{th}}$ percentile used in the main analysis. To ensure that the variation is driven by arbitrary differences in elevation rather than potentially endogenous decisions about precise canal placement, Panel C excludes settlements intersected by canals and Panel D adds an additional control variable for distance from the settlement to the nearest canal. Panel E restricts the sample to settlements proximate to canal segments that are long ($\geq 5km$) and straight (sinuosity ≤ 1.2), where we can be most confident that canal construction was not guided by efforts to include or exclude specific areas. Panel F shows results with the sample from the main analysis but with no land area weights to show robustness to our weighting choice. Panel G accounts for spatial correlation by estimating Conley standard errors (with a maximum distance for the spatial kernel of 100 km) with the main analysis sample. Table A9

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

estimates canal effects using the alternative command area boundary RDD described in Section 5.1, where distance to the command area boundary is the running variable rather than relative elevation.

The results are highly consistent across all of the specifications; major deviations from the main results that appear in more than one specification are noted here. Some specifications find evidence of substitution away from groundwater use in canal-irrigated areas; it is not surprising to find some substitution of this kind, but the magnitudes are small (<2 percentage points), especially relative to the increase in canal irrigation.[47] In the command area specification, we find higher *kharif* productivity effects than in the *rabi* season; in all other specifications, *rabi* effects are substantially higher (Panel B, Table A9). The non-ruggedness balanced sample specification (only) shows a small increase in the use of mechanized farm equipment (Panel A, Table A6). The null results on structural transformation are highly robust: we never estimate more than a 0.3 percentage point change in the non-farm employment share in any sector in any direction, though a handful of specifications show very small reductions in manufacturing or increases in services.[48] The population change effects are highly significant for all specifications. Importantly, our most restrictive specification, where we use only long and straight canal segments that do not show signs of local geographic targeting (Panel F), finds no meaningful differences from any of the results in our main specification. Table A10 further shows that these restricted sample results are consistent across a wide range of values for both the minimum canal length and maximum sinuosity.

Finally, we test for sensitivity of outcomes to different RDD parameter choices. Appendix Table A11 shows that treatment effects are highly stable in magnitude and significance across relative elevation bandwidths (Panel A), ruggedness balance restrictions (Panel B), and maximum distance to a canal (Panel C).

[47] The implications of our findings are similar even if there is some substitution away from groundwater — it would still imply an increase in agricultural productivity and a reduction in irrigation costs in canal-irrigated areas.

[48] Figure 3 puts the magnitude of these coefficients into perspective — the services coefficient in that figure represents a 0.3 percentage point change, the largest magnitude that we estimate.

6.2 Estimates of spillovers of canals to above-canal settlements

We next test for spillover effects in settlements that are close to canals, but at elevations just above them. The regression specification in Equation 5.2 generates separate estimates that compare these above-canal settlements to both below-canal (directly treated) and distant settlements, with matching on geophysical features.

Table 5 shows the results. The first row "Below-canal minus above-canal" is the difference between canal-treated settlements (as defined by relative elevation) and the omitted above-canal settlements. This is an alternate estimator of the direct effect of access to irrigation. The "Above-canal minus distant" coefficient is the coefficient of interest for studying spillovers. If canals affect the economy of *unirrigated* villages in the vicinity of the canal and spillovers decay across space, this coefficient will be different from zero.[49]

In the irrigation outcomes (Panel A), there are no substantive spillovers, confirming that our estimation is indeed isolating the direct effects of access to canal irrigation. Of particular note is the absence of effects on tubewell-irrigated area, indicating that groundwater recharge is not a major spillover channel for above-canal settlements. Similarly, there are few spillovers to above-canal settlements in agricultural productivity and land use (Panel B).[50]

Turning to non-farm outcomes, there are moderate spillovers on population density (Panel C); above-canal villages have about 5.2% higher density than otherwise-similar distant villages. This implies that the population density effect for below-canal (i.e. canal-irrigated) villages is also 5.2 percentage points higher than what we estimated previously, for a total effect on directly-irrigated villages of 20%.[51] Canals attract new rural residents not only directly to the irrigated areas, but to the periphery of those areas. These spillovers are both large in magnitude but compact. They substantially raise our estimates below of the net population flows caused by canals (Section 7),

[49] Note that the table reports the negative value of of γ_2 in Equation 5.2, such that a positive coefficient implies a positive spillover to settlements that are just above the canal.

[50] Water-intensive crops are grown more in above- *and* below-canal regions as compared with distant settlements. Note that this is not an acreage or volume measure, but a coarse indicator of whether a water-intensive crop is one of the three primary crops in the village.

[51] Note also that the "below minus above" coefficient is nearly identical to that in the RDD analysis earlier, as expected.

but they also demonstrate the value of being very close to irrigated land — net flows into areas 0–10km from the canal are only a quarter of the size of net flows into the irrigated zone itself.

In contrast, there is no evidence of spillovers in the measures of structural transformation, consumption, or education (Panels C and D). The coefficients on non-farm employment shares, and sectoral employment shares are all precisely-estimated zeroes. This analysis rules out the narrative of rural industrialization directly on the periphery of canals.[52]

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

The empirical strategies used thus far are best suited for measuring broad changes that occur across many settlements, and are either in the canal-irrigated space or in direct proximity to it. But if canals caused changes primarily in a small number of urban areas with market linkages to canals, the estimates above might not have the precision to capture such a concentrated effect. Further, while towns might appear in the vicinity of newly irrigated land, we would not necessarily expect them to appear exactly in the canal irrigation zone — irrigation of town land would provide no advantage as towns do not have meaningful agricultural sectors. Any town near enough to have market linkages with canal villages could be affected by their increased population and economic activity. The RDD approach is therefore less useful for studying town emergence.[53] Instead, we use the long panel of town populations to test whether town growth responds to nearby canal construction.

Table 6 shows difference-in-differences estimates from Equation 5.3 of the effect of canal construction on town size and appearance. Odd-numbered columns show the binary treatment with the De Chaisemartin and d'Haultfoeuille (2020) estimator, and even-numbered columns use the canonical difference-in-differences (TWFE) setup with a continuous treatment, which is the share of the town's 20 km radius catchment area that is in a canal command area. We focus on the binary treatment estimates below; the continuous treatment estimates are similar in effective magnitude.

[52]These results are robust to alternative specifications. Tables A12–A14 present spillover estimates using alternate distance thresholds (5km and 20km instead of 10km for the above-canal region) and alternate entropy balance inclusion parameters. In Table A15 we compare results for landed and landless households, similarly finding no evidence of spillovers to either group.

[53]We do test for the direct effect of canal irrigation on the likelihood that a settlement is a town, finding no effect (Table A3, Column 1).

Following nearby canal construction, towns are 10.3% larger in population (Panel A, Column 1) and grow 4.6% faster (Panel A, Column 3). The continuous treatment model finds similar results (Panel A, Columns 2 and 4). In Panel B, we test for town appearance at various population thresholds. Canal construction makes towns 3.2 percentage points more likely to appear, according to the Population Census definition of a town, which uses a population threshold of 5000. The remaining columns use higher thresholds; towns are more likely to first cross the 10,000 and 50,000 thresholds following canal construction, but there is no effect on the probability of crossing 100,000 or 500,000. This is unsurprising, as towns in excess of 50,000 typically have more diversified economies and will have their fortunes less closely tied to their proximate hinterlands. The continuous treatment effect results are similar — increasing the share of canal-irrigated land in a town's catchment by 50 percentage points generates about the same point estimates as the binary effect of crossing the 20% irrigation threshold.[54,55]

While we show large and robust effects of canals on regional urban population, we do not have the data to identify the mechanisms driving this growth. In our model, town growth comes from increased demand for non-agricultural goods from larger village populations and richer landowners, but other channels such as the capital channel studied by Bustos et al. (2020) are also possible. Consistent with the rural results, we find no evidence that canal-induced urban population growth involves a meaningful change in the structure of production: towns within 20 km of a canal, compared with towns further from canals, have similar rates of non-farm jobs per adult in 2013 (0.128 versus 0.132), and similar rates of manufacturing (0.024 versus 0.023), and services (0.091 versus 0.099) jobs per adult. In short, in both rural and urban India, canal construction increased local population without meaningfully changing sectoral compositions.

[54] The number of post-treatment observations in the panel is too small to empirically distinguish a functional form for the time path of the population change. In other words, it is difficult to measure whether canals affect urban population growth in perpetuity, or whether they result in a level change in population, which is converged to over several decades. We therefore consider both of these possibilities when we discuss the magnitude of these estimates in the next section.

[55] Appendix Table A16 shows that these results are robust to inclusion of state-year fixed effects, changing sample years, and using larger or smaller catchment area definitions.

6.4 Distinguishing Migration from Fertility and Mortality Change

We have shown that canal villages and canal-region towns have greater populations than they would in the absence of canals. In this section, we present suggestive evidence that the main driver of population change has been net migration, rather than differences in fertility or mortality.

While we cannot estimate settlement-level fertility and mortality in the canal-construction era, we can make inferences about past demographics by looking at the contemporary age structure of the population. First, we show that there are not *persistent* changes in fertility in canal villages: Appendix Table A3 uses the RDD to show that canal villages in fact have a marginally *lower* (0.2%) population aged 0–6. Second, we show that there is no evidence of past differential mortality, which we proxy by the 70+ population share, in the same table.[56] In short, we find no evidence of substantial mortality or fertility change in response to canal irrigation, but we are unable to rule out transitory changes long in the past, or changes with extremely broad spillover effects.

We next turn to migration, using India's 1987–88 National Sample Survey (NSS), which was collected toward the tail end of multiple decades of large-scale canal construction following India's independence in 1947. This is a district-level survey, so we cannot use the identification strategies from Sections 6.1 and 6.2, which require more geographically precise data. Instead, we examine whether districts where many canals were recently built had substantially more recent migrants.

Our outcome of interest is whether a person is an in-migrant to their current location, defined in the NSS as having had a past permanent place of residence that is different from their current one. We define the canal exposure variable as the share of the district's land which is in the command area of a canal built between 1951 and 1981. We control for state fixed effects and the the area of each district in a canal command area in 1951.

Panel A of Appendix Table A17 shows the results, which suggest that canals induced substantial in-migration into canal districts. The Column 2 estimate of 0.066 implies that a one standard deviation increase in canal coverage over the period 1951–1981 (12.4 percentage points) caused a

[56] We also did not find evidence of demographic bulges, which would suggest transitory fertility increases in the past, nor do we find evidence of changes in demographic structure in spillover specifications.

0.8 percentage point (3.4%) increase in the likelihood of being an in-migrant by 1987–88. Results are similar if we use a different base year (Columns 1 and 3). Reassuringly, in a placebo exercise regressing in-migration on the change in canal command area *after* the NSS survey (1991–2021), we find no such effects (Column 4).[57]

In Panel B, we run the same test, separately for migrants to and from urban and rural destinations. Inflows to canal districts are entirely driven by people coming from rural areas; they are moving to both rural and urban places within canal districts, consistent with the main findings in the paper.[58] The finding that most migrants are sourced from rural areas suggests that canal migration flows caused regional structural transformation, as many rural-to-urban migrants were likely shifting from agricultural to non-agricultural work given the much higher non-agricultural employment shares in urban India.

7 Discussion

The empirics and model combine to form a picture of how canals have reshaped India's economic geography. Canals created sharp spatial discontinuities in agricultural productivity. In irrigated villages, the return to land went up, growers shifted to more water-intensive crops, and demand for labor rose. Rising labor demand may have put upward pressure on wages in the short run, but in the long run, new workers were attracted to canal areas until wages were again equalized across space. In the new spatial equilibrium, canal-irrigated areas are more densely populated, but the returns to labor are no different from non-irrigated areas. In contrast, the returns to land—the fixed factor—remain higher in irrigated areas even decades after the canals were built.

Substantial structural transformation occurred, but new non-farm work opportunities were concentrated in cities. We think of these as production clusters whose agglomeration externalities and natural advantages make them superior locations for non-agricultural economic activity. The literature suggests a range of potential mechanisms that could drive the link between agricultural productivity gains and urban growth. Bustos et al. (2020) show that landowners in Brazil invested

[57]As noted above, post-1991 canal completion appears to be mostly rehabilitation rather than new canal construction. Either one could have effects on in-migration so it is supportive evidence to find that this placebo exercise yields no significant effects on in-migration in 1987-88 data.

[58]The table notes show additional details on variable construction and the specification used here.

land rents in urban areas that were connected by banking networks. Land rents could also be used to finance migration, another channel for urbanization and wealth accumulation among landowners (Clemens, 2014). Other sectoral linkages between greater agricultural output and non-farm industry are suggested by Johnston and Mellor (1961).[59]

How large were the population movements induced by canals? We can conduct a back-of-the-envelope calculation to understand the scale of the changes. We make several simplifying assumptions. First, we assume that our estimates are driven entirely by net population movement rather than by fertility or mortality.[60] Second, we need to transform the urban treatment effects from Table 6 into static changes in present-day urban populations. Column 1 suggests a static treatment effect of 10.3%, *i.e.* that canal towns are 10.3% larger than they would be in the absence of canals.[61] Third, following the heterogeneous town appearance results, we assume that these urban treatment effects apply to towns with populations less than 100,000. Finally, to estimate net rural population flows, we use the estimates from Table 5, multiplying the below-canal and above-canal treatment effects on population (22.0% and 5.2%, respectively) by the number of villages in below- and above-canal catchment areas in all of India.[62]

Under these assumptions, India's canals have drawn an additional 5 million people to cities and towns in canal regions, and an additional 48 million people to rural canal regions. Canals have thus created substantial changes in India's economic geography, with both spatial dimensions (represented by rural-to-rural movements) and sectoral dimensions (embodied in the rural-to-urban movements).

By studying the effects of irrigation at different geographic scales, our results can help to unify some of the findings in the prior literature. Foster and Rosenzweig (2004a) find that villages most exposed

[59]These mechanisms are difficult to estimate in our current setting, where we have time series data only on the urban population, but are an interesting subject for future work.

[60]While supported by the evidence in the prior section, this assumption serves only to simplify the exposition; the changes in the spatial distribution of the population are economically important whether driven by migration, fertility, or mortality.

[61]Alternately, we could use the growth estimate from Column 3 of the same table; if we assume that canal towns grew 4.6% faster per decade, and multiply by the median three decades since canal construction, we would find that canal region towns are 14.4% larger by 2011, resulting in slightly larger urban change estimates.

[62]The Table 5 estimates show that the population density in above-canal villages is 5.2% higher than in distant villages, and that the population density in below-canal (treated) villages is 16.0% higher than in above-canal villages. This implies that above-canal villages have 5.2% more population than if canals had not been built and below-canal villages have 22.0% higher population ($1.160*1.052=1.220$).

to the Green Revolution shifted their production structure *away* from industry and toward agriculture, a result reminiscent of the theoretical prediction of Matsuyama (1992) for an open economy. But their study is limited to villages; the industrialization that we measure is concentrated and occurs at some distance from the villages exposed to higher agricultural productivity. Bustos et al. (2016) found that the direction of structural change depended crucially on whether the technical change was labor-augmenting; the introduction of genetically-modified soy freed up labor to work in industry. Crucially, the units of observation in that paper are Brazilian municipalities, which have populations in the tens of thousands and incorporate the equivalents of Indian rural villages and urban towns. Our findings suggest that towns may be the key focal point for structural change when it occurs.

A limitation of our analysis is that, with limited data going back to the construction times of canals, measuring the aggregate effects of canals is difficult and beyond the scope of this paper. If labor was sufficiently mobile, then canals could have raised wages equally throughout the country, a result which would also generate a null relationship between access to canals and wages. Given the very large share of India's agricultural land that is irrigated by canals, we cannot rule out the possibility that labor-sending regions have also experienced higher wages as a result of the canal network. We therefore must remain agnostic on the nature of aggregate spillovers.

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

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

8 Conclusion

India's canal system provides a novel testing ground for examining the geographic relationship between agricultural productivity improvements and structural transformation. In the long run, we find that spatial equilibrium was restored primarily through substantial changes in the size of the landless population. Decades after canals were built, there are no differences in living standards between landless workers in canal and nearby non-canal settlements, and irrigated villages have similar non-farm activity to unirrigated villages. However, structural transformation has taken place, with towns emerging and growing disproportionately in canal regions.

The limitations of our work come from the absence of high-resolution longitudinal data to characterize the short run effects of canals and the mechanisms by which canals drove population growth. We provide suggestive evidence that canals induced large-scale migration into both rural and urban areas, and that these migrants came from rural areas. A deeper disentangling of the economic history through which India's canals dramatically shifted population and economic activity across space 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 labor migration. Most of India's canals were built in or before the License Raj era, when manufacturing investments were significantly inhibited by the state, and firms could not rapidly respond to changes in labor demand, potentially enhancing the role of mobile labor. Whether modern agricultural shocks will be equally equilibrated by labor flows remains an important question for future research.

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

References

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito, "General Equilibrium Effects in Space: Theory and Measurement," 2022. Working paper.

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

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

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

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

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

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

Ashraf, Quamrul and Oded Galor, "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 2011, *101* (5), 2003–2041.

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

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

Blakeslee, David, Aaditya Dar, Ram Fishman, Samreen Malik, Heitor Pelegrina, and Karan Singh, "Irrigation and the Spatial Pattern of Local Economic Development in India," *Journal of Development Economics*, 2023, *161.*

_ , 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," 2022. Working paper.

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

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

Cadell, A., *Settlement Report of the District of Muzaffarnagar: Including a Report on the Permanent Settlement of the Western Parganas of the District, and Also a Report on the Settlement of the Ganges Canal Tract*, North-Western Provinces and Oudh Government Press, 1873.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–96.

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

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

Comin, Diego, Danial Lashkari, and Martí Mestieri, "Structural Change with Long-run Income and Price Effects," *Econometrica*, 2021, *89* (1), 311–374.

Conley, Timothy G, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

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

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

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker, "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya," *Econometrica*, 2022, *90* (6), 2603–2643.

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

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

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

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

Fischer, Guenther, Freddy Nachtergaele, Sylvia Prieler, HT Van Velthuizen, Luc Verelst, and David Wiberg, "Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)," *IIASA, Laxenburg, Austria and FAO, Rome, Italy*, 2008, *10.*

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

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

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

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

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

Funk, Chris C., Pete J. Peterson, Martin F. Landsfeld, Diego H. Pedreros, James P. Verdin, James D. Rowland, Bo E. Romero, Gregory J. Husak, Joel C. Michaelsen, and Andrew P. Verdin, "A Quasi-global Precipitation Time Series for Drought Monitoring," *Data Series*, 2014.

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

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

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

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

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

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

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

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

Hanlon, W Walker and Stephan Heblich, "History and Urban Economics," *Regional Science and Urban Economics*, 2022, *94*.

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.

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," *American Economic Review*, 2022, *112* (7), 2316–52.

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

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

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

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

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

Liu, Maggie, Yogita Shamdasani, and Vis Taraz, "Climate Change and Labor Reallocation: Evidence from Six Decades of the Indian Census," *American Economic Journal: Economic Policy*, 2023, *15* (2), 395–423.

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.

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2004, *72* (1), 159–217.

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

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

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

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

Pant, Niranjan, *Some Aspects of Irrigation Administration: A Case Study of Kosi Project*, Arunima Print Works, 1981.

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

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

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

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

Santangelo, Gabriella, "Firms and Farms: The Impact of Agricultural Productivity on the Local Indian Economy," 2019. Working paper.

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

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

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

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

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

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

_ , David Molden, Ramaswamy Sakthivadivel, and David Seckler, "Global Groundwater Situation: Opportunities and Challenges," *Economic and Political Weekly*, 2001, pp. 4142–4150.

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.

Stone, Ian, *Canal Irrigation in British India: Perspectives on Technological Change in a Peasant Society*, Cambridge University Press, 1984.

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

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

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

Table 1: Summary statistics

	All India	All canal-area settlements	All canal-area settlements minus donut hole	Ruggedness-balanced analysis sample
Sample Size	589,950	227,416	124,001	84,868
Percent Treatment	–	83	77	79
Means				
Total irrigated area (share of ag. land)	0.464	0.584	0.515	0.534
Canal irrigated area (share of ag. land)	0.134	0.185	0.144	0.137
Tubewell irrigated area (share of ag. land)	0.196	0.264	0.225	0.241
Other irrigated area (share of ag. land)	0.142	0.149	0.166	0.170
Agricultural land (share of total village area)	0.577	0.664	0.623	0.639
Kharif agricultural production, EVI-derived (log)	7.560	7.745	7.715	7.695
Rabi agricultural production, EVI-derived (log)	7.231	7.375	7.294	7.292
Any water intensive crop grown	0.586	0.648	0.594	0.593
Mechanized farming equipment (share of households)	0.047	0.063	0.055	0.061
Population density (log)	5.065	5.674	5.483	5.515
Consumption (log)	9.726	9.757	9.749	9.760
Total non-farm employment (share of adult pop)	0.096	0.086	0.088	0.086
Services employment (share of adult pop)	0.066	0.058	0.059	0.059
Manufacturing employment (share of adult pop)	0.019	0.020	0.020	0.020
Primary school ed attained (share of adult pop)	0.471	0.498	0.487	0.495
Middle school ed attained (share of adult pop)	0.318	0.339	0.327	0.330
Secondary school ed attained (share of adult pop)	0.194	0.211	0.205	0.206
Literacy rate (literate share of adult pop)	0.561	0.578	0.576	0.580

Notes: This table shows summary statistics for the main outcomes in various samples of the data. The All India sample includes every village or town recorded in the 2011 Population Census with a non-zero population. The all canal-area settlements sample includes towns and villages $\leq$10km from the nearest canal, and within 50m of the nearest canal in terms of elevation. In the third column, removing the donut hole from all canal-area settlements drops settlements $\pm$2.5m in elevation from the nearest canal from the sample. We then impose a balance criteria on ruggedness by dropping settlements from subdistricts in which there is a $\geq$25% difference in average ruggedness between below-canal (treatment) and above-canal (control) settlements. The resulting sample, with 84,868 settlements, is the ruggedness-balanced analysis sample and is our preferred sample used in the RDD analysis. Note that the mean values reported for the ruggedness-balanced analysis sample also exclude subdistricts that do not contain at least one settlement in each of the treatment and control groups. All mean values are weighted by land area.

Table 2: Balance in the regression discontinuity design

	Ruggedness (TRI)	Annual rainfall avg. 2010-2014 (mm)	Max monthly temp. avg. 2010-2014 (°C)	Soil quality
Below canal	0.053	-0.402	0.037***	0.005
	(0.068)	(1.576)	(0.008)	(0.007)
Control group mean	4.809	1049.216	32.540	0.841
Observations	84,763	84,763	84,763	84,763
R^2	0.63	0.99	0.98	0.55

	Distance to coast (km)	Distance to river (km)	Wetland rice (GAEZ)	Wheat (GAEZ)
Below canal	-0.177	-1.481***	0.000	0.000
	(0.387)	(0.341)	(0.012)	(0.004)
Control group mean	328.402	24.293	2.119	0.547
Observations	84,763	84,763	84,763	84,763
R^2	1.00	0.88	0.93	0.98

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

Notes: This table reports the regression discontinuity estimates for geophysical variables following Equation 5.1, dropping each outcome variable from the list of controls for each result. The Terrain Ruggedness Index (TRI) is a topographic measure of ruggedness, essentially a measure of the variance of elevation in a settlement area. Rainfall is calculated as the average total annual rainfall in a settlement, measured over 2010-2014. Temperature is measured as the average maximum monthly temperature from 2010-2014. The soil quality is a binary variable where a value of 1 indicates no limitations for rooting conditions and value of 0 indicates moderate to severe limitations. Rooting conditions are modeled based on the soil depth, volume, and presence of gravel. The Global Agro-Ecological Zones (GAEZ) model estimates predicted yield to give us a measure of crop suitability for wetland rice and wheat in each settlement. The model is based on expected conditions for agricultural production such as climate, soil, and terrain parameters and assumes gravity-fed irrigation and intermediate level inputs.

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	0.075***	0.099***	-0.011*	-0.004
	(0.008)	(0.007)	(0.007)	(0.005)
Control group mean	0.428	0.032	0.213	0.189
Observations	76,618	76,622	76,678	75,888
R^2	0.61	0.38	0.47	0.64

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Below canal	0.027***	0.017*	0.071***	0.027***	0.002
	(0.006)	(0.009)	(0.012)	(0.009)	(0.002)
Control group mean	0.595	7.692	7.210	0.555	0.057
Observations	83,512	83,450	83,190	65,691	79,972
R^2	0.61	0.83	0.71	0.72	0.31

Panel C. Non-farm outcomes

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log)
Below canal	0.154***	0.001	0.003*	-0.001	0.007
	(0.028)	(0.002)	(0.001)	(0.001)	(0.006)
Control group mean	5.239	0.090	0.059	0.020	9.743
Observations	84,763	79,291	79,291	79,291	80,677
R^2	0.42	0.26	0.19	0.28	0.52

Panel D. Education outcomes

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (literate share of pop.)
Below canal	0.013***	0.013***	0.010***	0.011***
	(0.004)	(0.003)	(0.002)	(0.002)
Control group mean	0.476	0.311	0.196	0.569
Observations	79,924	79,924	79,924	84,763
R^2	0.56	0.55	0.52	0.57

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

Notes: This table reports the main regression discontinuity estimates following Equation 5.1 for all outcomes variables. Each outcome variable is estimated separately, with the β_1 coefficient capturing the direct effect of canal irrigation reported in the first row of each panel. Panel A reports effects on the share of agricultural land irrigated by various sources. Panel B reports effects on agricultural outcomes, including the satellite-derived proxies agricultural productivity by season. Panel C reports non-farm outcomes, including population density, employment rates by sector, and the log of per capita consumption. Panel D reports education outcomes. For each reported coefficient, stars indicate the coefficient's significance and the standard error is reported below in parentheses. The control group mean (weighted by land area), the number of observations with non-missing data for the particular outcome variable, and the adjusted R^2 for each regression estimate are also shown.

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

Panel A. Land ownership overview

	Land-owning HHs (share of all HHs)	Avg. size of land holdings (log hectares, all HHs)	Avg. size of land holdings (log hectares, land-owning HHs)
Below canal	-0.027***	-0.055***	0.006
	(0.005)	(0.019)	(0.014)
Control group mean	0.534	0.745	1.525
Observations	79,972	77,756	77,723
R^2	0.46	0.46	0.50

Panel B. Consumption distribution

	Consumption pc		Consumption pc (log, land-owning HHs)			
	(log, landless HHs)	(log, land-owning HHs)	1^{st} quartile	2^{nd} quartile	3^{rd} quartile	4^{th} quartile
Below canal	0.002	0.021***	0.000	0.015**	0.019***	0.029***
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.007)
Control group mean	9.603	9.812	9.737	9.763	9.810	9.904
Observations	77,791	77,720	67,968	71,126	71,860	69,404
R^2	0.46	0.55	0.45	0.46	0.45	0.41

Panel C. Education attainment

	At least primary, share of		At least middle, share of		At least secondary, share of	
	landless pop.	land-owning pop.	landless pop.	land-owning pop.	landless pop.	land-owning pop.
Below canal	0.011***	0.022***	0.011***	0.023***	0.007***	0.019***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Control group mean	0.431	0.515	0.268	0.351	0.160	0.231
Observations	77,638	78,018	77,638	78,018	77,638	78,018
R^2	0.46	0.59	0.45	0.57	0.41	0.54

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

Notes: This table reports the regression discontinuity estimates following Equation 5.1 for additional outcomes pertaining to land ownership. Each outcome variable is estimated separately, with the β_1 coefficient capturing the direct effect of canal irrigation reported in the first row of each panel. Panel A shows estimates for the share of households that are landowners, the log of average size of land holdings for all households, and the log of average size of land holdings among only land-owning households. Panel B first shows estimates for consumption disaggregated by land ownership status. Panel B then shows estimates for consumption by nationally-defined landholding quartiles, with each quartile of land-owning households separately estimated. The bottom (1^{st}) quartile are the land-owning households with total land holdings in the 0–25% range of the national distribution while the top (4^{th}) quartile are those in the top 75–100%. The first quartile owns 0–1 hectare of land, the second owns 1–2 hectares, the third owns 2–4 hectares, and the fourth owns more than 4 hectares. Note that all consumption coefficients are in units of log consumption per capita, as they are throughout the paper.

Table 5: Comparison to distant settlements

Panel A. Irrigation outcomes

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Below-canal minus	0.046***	0.083***	-0.009	-0.017**
above-canal settlements	(0.013)	(0.010)	(0.008)	(0.008)
Above-canal minus	0.009	0.005	0.007	-0.003
distant settlements	(0.007)	(0.004)	(0.006)	(0.005)
Control group mean	0.450	0.068	0.212	0.177
Observations	76,014	76,196	76,185	75,569
R^2	0.62	0.17	0.42	0.79

Panel B. Agriculture outcomes

Settlement type	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Below-canal minus	0.020***	0.015	0.055***	0.005	0.007***
above-canal settlements	(0.007)	(0.011)	(0.020)	(0.012)	(0.002)
Above-canal minus	-0.004	0.003	-0.028	0.043**	0.000
distant settlements	(0.007)	(0.009)	(0.018)	(0.017)	(0.002)
Control group mean	0.572	7.821	7.337	0.659	0.038
Observations	84,682	84,654	84,467	63,937	80,887
R^2	0.56	0.87	0.58	0.71	0.32

Panel C. Non-farm outcomes

Settlement type	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Below-canal minus	0.160***	0.001	0.002	0.000	0.024***
above-canal settlements	(0.028)	(0.002)	(0.001)	(0.001)	(0.006)
Above-canal minus	0.052**	-0.001	0.001	-0.001	0.003
distant settlements	(0.024)	(0.003)	(0.001)	(0.002)	(0.007)
Control group mean	5.634	0.083	0.053	0.021	9.637
Observations	85,762	78,572	78,572	78,572	81,351
R^2	0.27	0.14	0.09	0.22	0.43

Panel D. Outcomes disaggregated by land ownership

Settlement type	Consumption pc (log, landless HHs)	Consumption pc (log) (log, land-owning HHs)	Middle school ed. (share of landless pop.)	Middle school ed. (share of land-owning pop.)
Below-canal minus	0.008*	0.023***	0.011***	0.023***
above-canal settlements	(0.004)	(0.006)	(0.003)	(0.005)
Above-canal minus	-0.009	0.008	-0.001	0.004
distant settlements	(0.008)	(0.009)	(0.004)	(0.006)
Control group mean	9.502	9.739	0.256	0.359
Observations	78,325	78,683	78,142	78,852
R^2	0.36	0.43	0.44	0.54

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

Notes: This table reports the spillover analysis estimates following Equation 5.2, comparing the below-canal (directly treated) and distant settlements to the above-canal settlements (the omitted group). The below-canal and above-canal settlements are defined in the same way the main RDD sample was defined (the ruggedness-balanced analysis sample shown in Table 1). Distant settlements are defined as being 15–50km from the canal and at a higher elevation than the nearest canal so that it is not feasible for them to receive canal irrigation. Weights were calculating using entropy balancing to ensure distant settlements are comparable to above-canal settlements with respect to all geophysical controls following Hainmueller (2012). The γ_1 (below-canal minus above-canal settlements) and $-\gamma_2$ (above-canal minus distant settlements) estimates are reported here. We report $-\gamma_2$ rather than γ_2 so that positive coefficient values reflect positive spillovers into the above-canal settlements to ease interpretation. The control group means reflect the area-weighted mean values of the above-canal settlements. Standard errors are clustered at the district level.

Table 6: Effect of canals on town size and population

Panel A. Town population and growth

	Log Population		Log Pop Growth	
Command area in town catchment area	0.103***		0.046**	
(binary treatment)	(0.031)		(0.023)	
Share of town catchment area in command area		0.263***		0.063**
(continuous treatment)		(0.043)		(0.028)
Observations	302691	64260	263628	58905
R^2		0.830		0.150

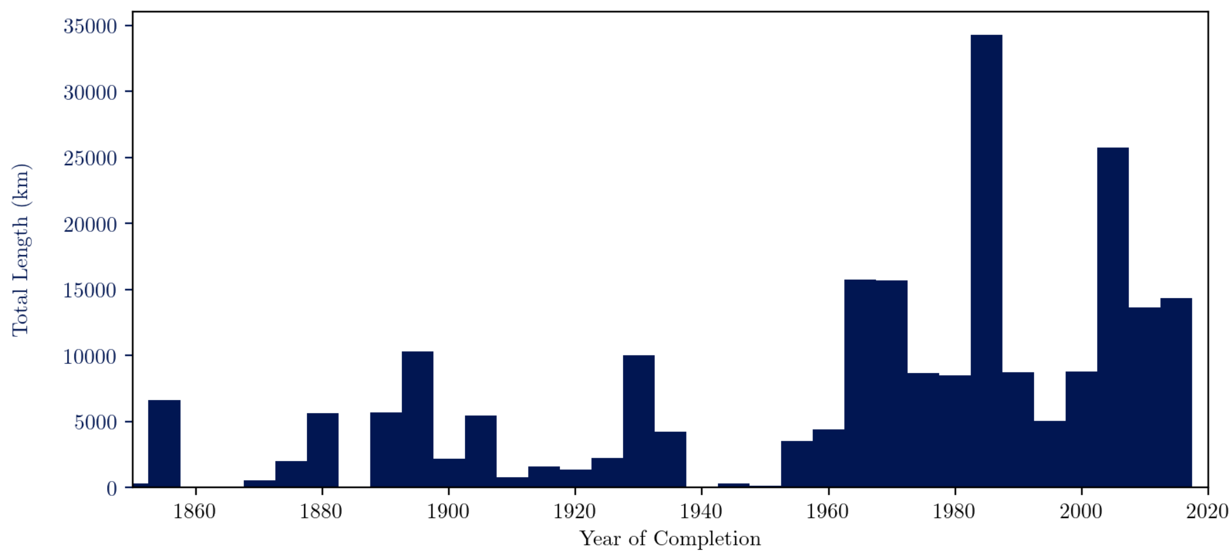
Panel B. Town appearance

	Pop. 5000		Pop. 10,000		Pop. 50,000		Pop. 100,000		Pop. 500,000	
Command area in town catchment area	0.032***		0.041***		0.015**		0.005		-0.001	
(binary treatment)	(0.013)		(0.016)		(0.007)		(0.004)		(0.001)	
Share of town catchment area in command area		0.079***		0.101***		0.040***		0.016*		-0.004
(continuous treatment)		(0.018)		(0.021)		(0.012)		(0.009)		(0.002)
Observations	302691	64260	302691	64260	302691	64260	302691	64260	302691	64260
R^2		0.700		0.650		0.520		0.470		0.350

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

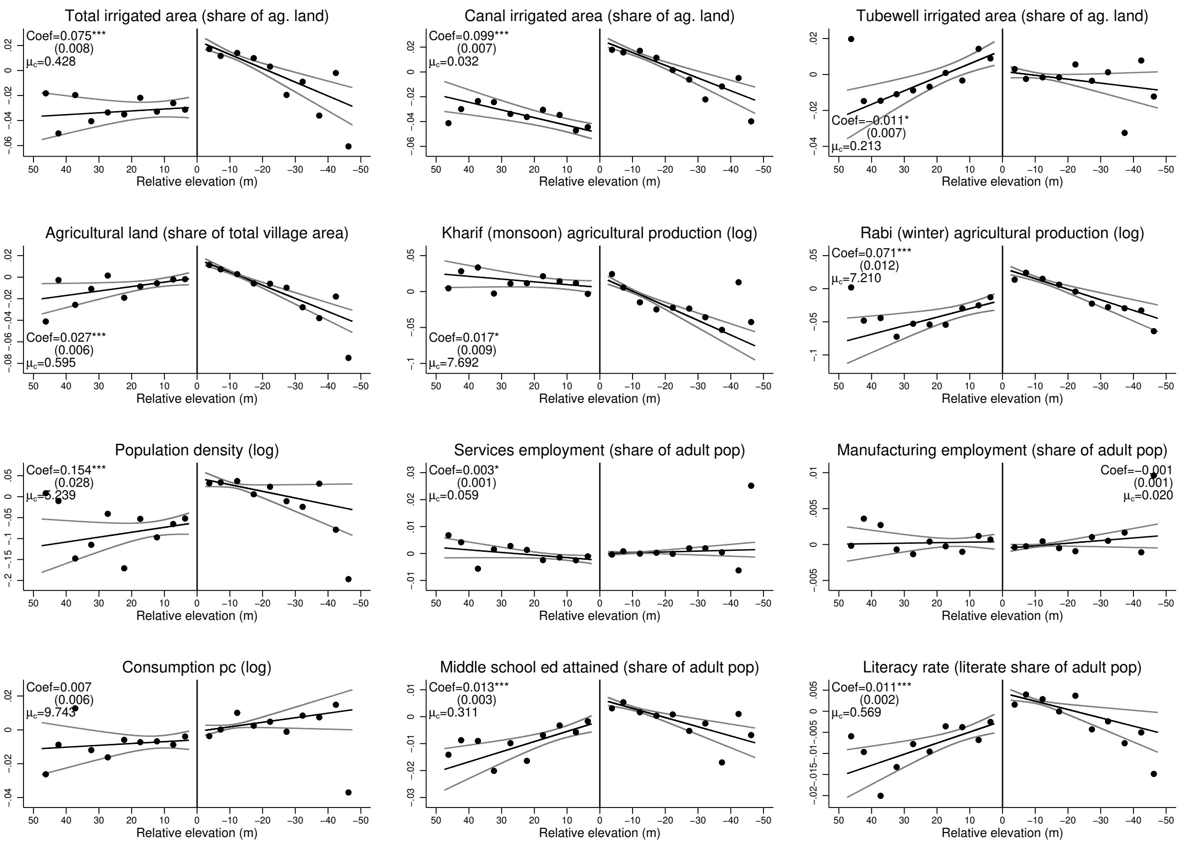
Notes: This table shows effect of canal construction on town size and population, as identified by the β_1 in Equation 5.3. The outcome variable in columns 1 and 2 is log town population. Before a town appears in the time series, we assign it a population of 2,000, reflecting the typical size of settlements before they become towns. In subsequent columns, the outcome variable is an indicator that takes the value of one once a town has exceeded a certain population threshold. This indicator is set to 0 for the decades before a town appears in the census data. Odd-numbered columns define canal construction with an indicator that takes the value 1 once 20% of the town's catchment area (a circle with 20 km radius) has been covered by a command area. These estimates are calculated using De Chaisemartin and d'Haultfoeuille (2020). Even-numbered columns show results from standard two-way fixed effect (TWFE) continuous treatment regressions, where we show the coefficient on the share of the town catchment area covered by a command area.

Figure 1: Canal construction through time



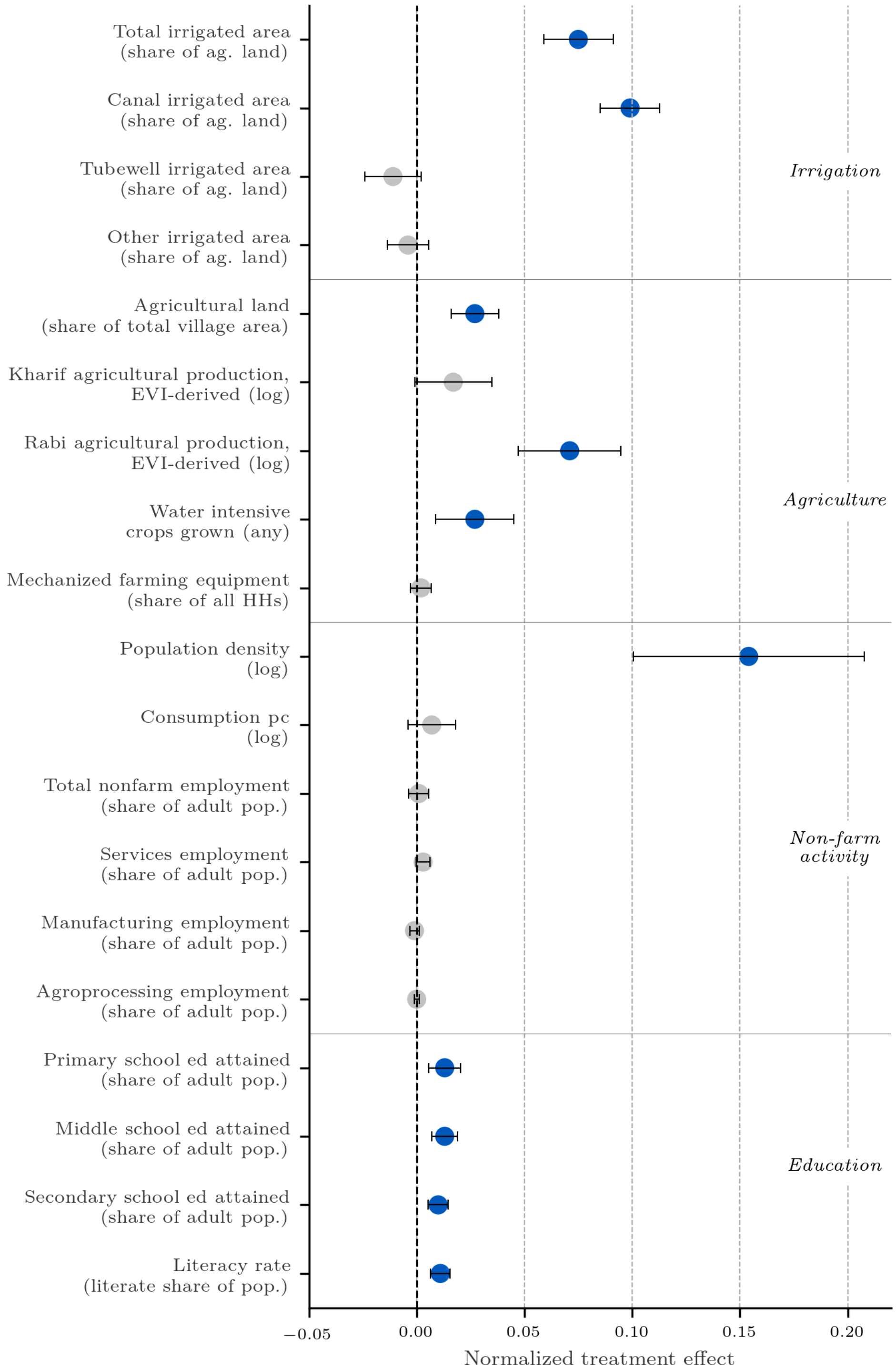
Notes: This plot shows the total length of medium and major canals constructed in India from 1850–2013. Any canals with dates older than 1850 are coded as 1850 while any canals not completed before 2013 are not included. Note that 150 of the 1442 total canal projects reported, or 6% of total canal length in the geospatial canals data, have an unknown date of completion and are not included in this plot. Additionally, 313 projects totaling 26% of total canal length in the data were not completed as of 2013 (the last date of our major outcomes) and so are not included in this plot.

Figure 2: Regression discontinuity binscatters for key outcomes



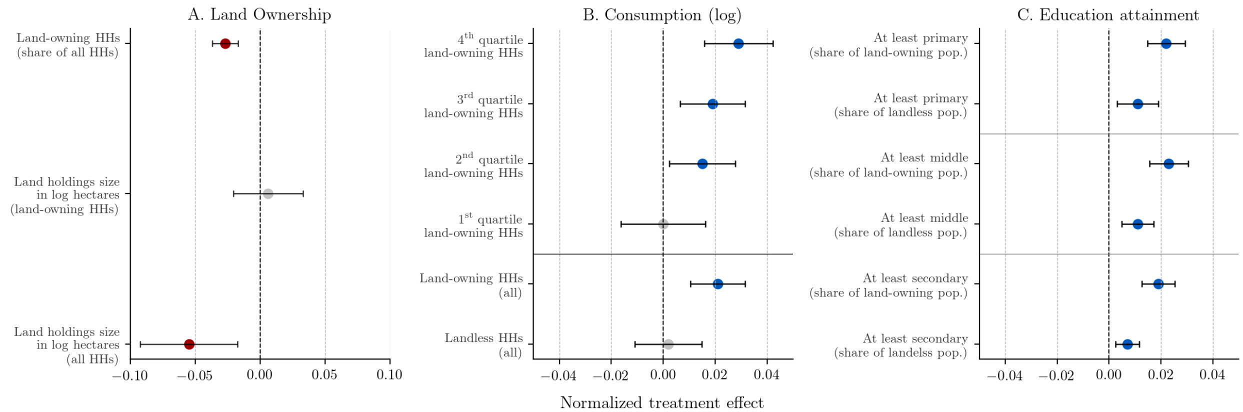
Notes: Each figure shows the binned scatterplot relationship between an outcome of interest and the RDD running variable (elevation relative to the nearest canal), after residualizing on the geophysical controls and subdistrict fixed effects. Below-canal (directly treated) settlements have negative relative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

Figure 3: Regression discontinuity results for main outcomes



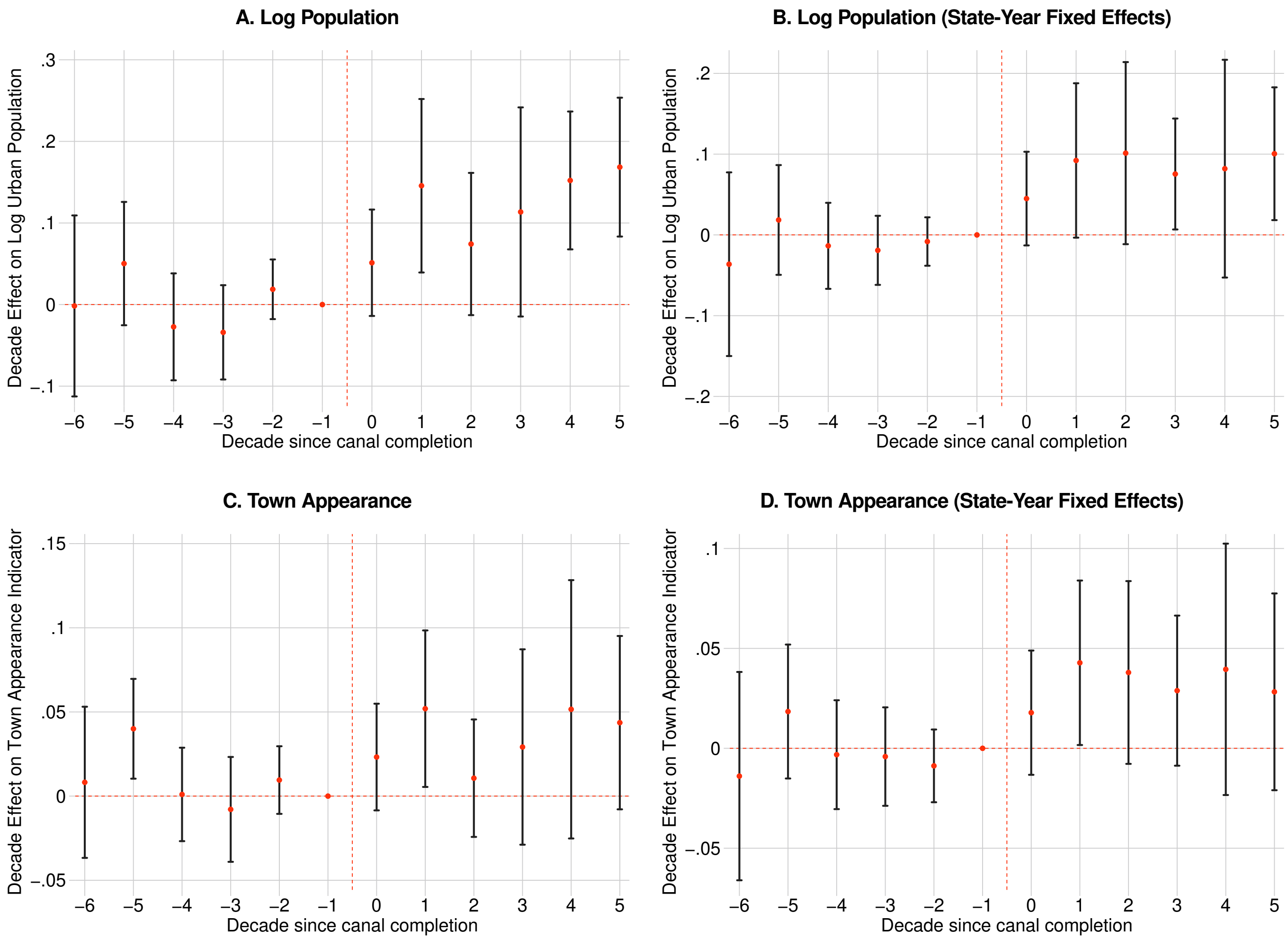
Notes: This figure shows the normalized β_1 regression discontinuity estimates for the main outcomes variables following Equation 5.1 and reported in Table 3. Blue points indicate positive, significant normalized treatment effects while gray points indicate results not significant at the 95% level. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in control settlements of the analysis sample. Error bars indicate the 95% confidence interval for each estimate.

Figure 4: Land ownership outcomes



Notes: This figure shows the normalized β_1 regression discontinuity estimates for various outcomes pertaining to land ownership, following Equation 5.1 and reported in Table 4. Blue points indicate positive, significant normalized treatment effects, red points indicate negative, significant normalized treatment effects, and gray points indicate results not significant at the 95% level. The normalized treatment effect is calculated by dividing the regression discontinuity coefficient by the standard deviation of the outcome variable in the control settlements of the analysis sample. Error bars indicate the 95% confidence interval for each estimate.

Figure 5: Difference-in-differences estimates of effects of canal construction on town appearance and size



Notes: The figure shows difference-in-differences plots, calculated following De Chaisemartin and d'Haultfoeuille (2020) which describe the effect of canal construction on urban population (Panels A and B) and town emergence (Panels C and D). Each point shows a regression estimate describing the relative value of the outcome variable x decades after canal completion. Year 0 is the first census year following canal completion. A town is considered treated by a canal in the first decade when 20% of the 20km radius circle around the town is in a canal's command area. All estimates control for town and decade fixed effects; town populations are observed every decade. Town appearance is an indicator that takes the value one in any census year where the town is observed with population over 5000. For the population regressions, towns that have not yet appeared in the census are assigned a population of 2000. Standard errors are clustered at the district level.

A Appendix Tables and Figures

Table A1: Variable dictionary

Population Census 2011

Agricultural land	The share of total village area used for agriculture. Total area is defined by the extent of the village boundary in the GIS data while agricultural land area is reported in the census in hectares.
Total irrigated area	The share of agricultural land in the village that is irrigated by any method. Both total agricultural land area and total irrigated land area are reported in the census in hectares.
Canal irrigated area	The share of agricultural land in the village that is irrigated by canals. Both total agricultural land area and canal-irrigated land area are reported in the census in hectares.
Tubewell irrigated area	The share of agricultural land in the village that is irrigated by tubewells. Both total agricultural land area and tubewell-irrigated land area are reported in the census in hectares.
Other irrigated area	The share of agricultural land in the village that is irrigated by other methods such as lakes or tanks. Both total agricultural land area and other-irrigated land area are reported in the census in hectares.
Water intensive crops	A binary indicator where 1 indicates that the village reports growing cotton, sugarcane, or rice while 0 indicates the village does not report growing any of the those three crops. Each village reports the top three crops grown in the village, which comprises the list this variable is created from.
Mechanized farm equipment	The share of households in the village who own mechanized farming equipment.
Population density	The total settlement (village or town) population per square km, with settlement area defined using the settlement boundary from the GIS data.
Literacy rate	The share of the settlement population that is literate.

Economic Census 2013

Total employment	The percentage of the adult population employed in non-farm work in each settlement. Non-farm employment is the sum of employees reported by firms located in each settlement. Adult population is defined as all people aged 18 years and older from the 2011 population census. The top 1% outliers are top coded with the 99th percentile value.
Services employment	The percentage of the adult population employed in services in each settlement. Services employment is the sum of employees reported by firms corresponding to NIC codes 36-93 and 131 located in each settlement. Adult population is defined as all people aged 18 years and older from the 2011 population census. The top 1% outliers are top coded with the 99th percentile value.
Manufacturing employment	The percentage of the adult population employed in manufacturing in each settlement. Manufacturing employment is the sum of employees reported by firms corresponding to NIC codes 10-35 (excluding only 131) located in each settlement. Adult population is defined as all people aged 18 years and older from the 2011 population census. The top 1% outliers are top coded with the 99th percentile value.
Agroprocessing employment	The percentage of the adult population employed in agroprocessing in each settlement. Agroprocessing employment is the sum of employees reported by firms corresponding to NIC codes 101-110, excluding 109, and 120 located in each settlement. Adult population is defined as all people aged 18 years and older from the 2011 population census. The top 1% outliers are top coded with the 99th percentile value.

Socioeconomic Caste Census (SECC) 2012

Predicted consumption per capita	Consumption per capita in each settlement. Household consumption is predicted from a list of household assets, following the methodology of Elbers et al. (2003) and described in detail in Asher and Novosad (2020). Household consumption is summed across households in a village and divided by the SECC-reported total population of that village. We then take the natural log to create the consumption per capita measure used throughout the paper.
Land-owning households	The share of households in a settlement that report owning non-zero land area.
Land holding size	The log of the average land holding size across all settlements in a village. Land holdings are originally reported in hectares.

Consumption by landholding quartiles	The log of the average predicted household consumption for land-owning households within each landholding quartile in a settlement. Consumption is estimated from the household asset list as described above. Landholding quartiles are defined by the national distribution. Households in the first quartile own 0–1 hectare of land, 1–2 hectares in the second, 2–4 hectares in the third, and households in the fourth own more than 4 hectares.
At least primary education	The share of the adult population in a settlement that has attained at least a primary school education. Level of education and age are reported at the individual level in the household roster in the SECC. The adult population is defined as individuals 18 years or older from the SECC.
At least middle education	The share of the adult population in a settlement that has attained at least a middle school education. Level of education and age are reported at the individual level in the household roster in the SECC. The adult population is defined as individuals 18 years or older from the SECC.
At least secondary education	The share of the adult population in a settlement that has attained at least a high school education. Level of education and age are reported at the individual level in the household roster in the SECC. The adult population is defined as individuals 18 years or older from the SECC.

Enhanced Vegetation Index (EVI), 2011-2013

Kharif agricultural production	A measure of green-up for the May-October growing season. The average EVI value from the first six weeks of the Kharif season is subtracted from the maximum EVI value achieved in each settlement during the season. The log of this difference is our proxy measure for Kharif productivity. Note that EVI is calculated at the pixel level using raster data, then extracted to the settlement level using the settlement boundaries.
Rabi agricultural production	A measure of green-up for the late December through late March dry season. The average EVI value from the first six weeks of the Rabi season is subtracted from the maximum EVI value achieved in each settlement during the season. The log of this difference is our proxy measure for Rabi productivity. Note that EVI is calculated at the pixel level using raster data, then extracted to the settlement level using the settlement boundaries.

Shuttle Radar Topography Mission (SRTM)

Relative elevation	SRTM reports global elevation in a 90m-resolution grid, from which we extract the full distribution of pixels that lie within a settlement boundary to characterize the settlement's elevation. The relative elevation of a settlement is calculated as the difference between the 5th percentile of elevation in the settlement and the elevation on the nearest point of the nearest canal.
Ruggedness	The terrain ruggedness index (TRI) measures ruggedness as the average square difference in elevation between a pixel and its eight surrounding pixels. We calculate the TRI value for every pixel in the elevation gird, then take the average TRI value across all pixels within a settlement's boundary to calculate that settlement's ruggedness.

Climate Hazards Center, 2011-2013

Annual rainfall	A settlement's average total annual rainfall from 2010–2014. The Climate Hazards Center InfraRed Precipitation with Station (CHIRPS) dataset (Funk et al., 2014) reports gridded, 10-day rainfall. Using the settlement boundaries, we sum total annual rainfall for the settlement for each year from 2010–2014. We then calculate the average annual rainfall over that time period.
Maximum monthly temperature	A settlement's average maximum monthly temperature from 2010–2014. The Climate Hazards Center Infrared Temperature with Stations (CHIRTS) dataset (Funk et al., 2019) dataset reports gridded, maximum daily temperature. Using the settlement boundaries, we take the maximum temperature for each month in each settlement from 2010–2014. We then calculate the average maximum monthly temperature over that time period.

FAO Global Agro-Ecological Zones (GAEZ)

Wetland rice	Estimated yield of wetland rice for each settlement as calculated by the FAO GAEZ model assuming intermediate level inputs and gravity-fed irrigation. The model predicts gridded yields based on climate, soil, and terrain parameters. We extract the values within a settlement's boundary to calculate the mean estimated yield for each settlement as a measure of suitability for wheat production.

Wheat	Estimated yield of wheat for each settlement as calculated by the FAO GAEZ model assuming intermediate level inputs and gravity-fed irrigation. The model predicts gridded yields based on climate, soil, and terrain parameters. We extract the values within a settlement's boundary to calculate the mean estimated yield for each settlement as a measure of suitability for wheat production.

Harmonized World Soil Database (HWSD)

Soil quality	A binary variable where 1 indicates no limitations for rooting conditions and 0 indicates moderate or severe limitations. HWSD models this measure of rooting conditions based on soil depth, volume, and presence of gravel. Because the model is categorical, we define a settlement's value as the most frequently occurring value by pixel within the settlement boundary.

National Sample Survey (NSS) 1987–88 Migration Module

Is a migrant	A binary variable where 1 indicates the respondent reported having a different previous place of residence than the current residence at the time of enumeration. Note that the NSS is designed to be representative at the district level.
Is a migrant from a rural area	A binary variable where 1 indicates a migrant (as defined above) who reported their previous place of residence in an urban area, and 0 is all other respondents (including non-migrants).
Is a migrant from an urban area	A binary variable where 1 indicates a migrant (as defined above) who reported their previous place of residence in an urban area, and 0 is all other respondents (including non-migrants).

Water Resources Information System (WRIS): Canals and Command Areas

Share of town catchment area in command area	The share of the 20km-radius circle drawn around each town (defining the catchment area of the town) that overlaps with a canal's command area. We draw each town's catchment area around the town's point then calculate the spatial overlap with the command area polygons provided by WRIS. (This is a town-level measure used in the town population panel analysis).
Command area in town catchment area	A binary variable where 1 indicates that $\geq 20\%$ of the town's catchment area is covered by a command area, and 0 indicates $< 20\%$ coverage. (Town-level measured used in the town population panel analysis).

Canal coverage	The share of a district area that is covered by a canal's command areas. We calculate the percent overlap between the district polygons and the command area polygons from WRIS in GIS. (This is a district-level measure used in the NSS migration analysis).
Canal coverage gain	The percentage point difference the canal coverage (as defined above) between two points in time. (This is a district-level measure used in the NSS migration analysis),
Inside command area	A binary variable where 1 indicates that a settlement's centroid is located inside a command area, and 0 indicates that the town's centroid is outside of a command area.

Geospatial Measures (GIS)

Distance to canal	The straightline distance (in kilometers) from a settlement's centroid to the nearest point on the nearest canal.
Distance to coast	The straightline distance (in kilometers) from a settlement's centroid to the nearest point of coastline.
Distance to river	The straightline distance (in kilometers) from a settlement's centroid to the nearest point on the nearest river.

Table A2: Balance in the regression discontinuity design (1951 Population Census characteristics)

	Population	Sex ratio	Population density (log)	HH size	Literacy rate
Below canal	7.267	0.090	-0.249	-0.363	0.033
	(78.010)	(0.215)	(0.353)	(0.335)	(0.067)
Control group mean	570.897	1.492	-4.816	4.818	0.338
Observations	4,172	4,039	820	767	402
R^2	0.24	0.22	0.36	0.31	0.24

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

Notes: This table reports the regression discontinuity estimates for 1951 Population Census village characteristics following Equation 5.1. Population is the total population in the village. Sex ratio is the number of males divided by number of females. Population density is total population divided by village area (in square miles). HH size is the mean household size, generated by dividing the total population by the number of occupied houses. Literacy rate is the number of literate people in the village divided by total population. We were able to extract and match 1951 data for 32,765 villages in 109 districts across six states (Gujarat, Karnataka, Madhya Pradesh, Maharashtra, Rajasthan, and Uttar Pradesh). The sample in each column is the main analysis sample matched to the 1951 village data we were able to parse whose nearest canal was completed after 1951 and for which the outcome data was available in our matched sample.

Table A3: Regression discontinuity results for additional outcomes

	Settlement is a town (likelihood)	Population age 0-6 (share of pop.)	Population age 70+ (share of pop.)	Agroprocessing emp. (share of adult pop.)
Below canal	0.009	-0.002***	0.000	0.000
	(0.007)	(0.001)	(0.000)	(0.000)
Control group mean	0.024	0.140	0.036	0.006
Observations	84,763	84,763	79,966	79,291
R^2	0.14	0.57	0.32	0.43

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

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

Table A4: Balance in the regression discontinuity design using distance to command area boundary

	Ruggedness (TRI)	Annual rainfall avg. 2010-2014 (mm)	Max monthly temp. avg. 2010-2014 (°C)	Soil quality
Inside command area	-0.011	0.648	0.026***	-0.005
	(0.042)	(3.324)	(0.010)	(0.009)
Control group mean	3.803	1181.072	32.452	0.909
Observations	48,809	48,809	48,809	48,809
R^2	0.65	0.99	0.99	0.77
	Distance to coast (km)	**Distance to river (km)**	**Wetland rice (GAEZ)**	**Wheat (GAEZ)**
Outside command area	0.128	-0.716	-0.002	0.007
	(0.457)	(0.761)	(0.016)	(0.006)
Control group mean	447.096	26.275	2.457	0.944
Observations	48,809	48,809	48,809	48,809
R^2	1.00	0.93	0.96	0.99

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

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

Table A5: Regression discontinuity results for irrigation outcomes (robustness)

	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, minus donut hole				
Below canal	0.079***	0.111***	-0.011**	-0.009**
	(0.007)	(0.006)	(0.005)	(0.004)
Control group mean	0.412	0.035	0.199	0.183
Observations	113,428	113,475	113,545	112,057
R^2	0.59	0.39	0.48	0.63
Panel B: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation				
Below canal	0.074***	0.107***	-0.017***	-0.007
	(0.007)	(0.006)	(0.005)	(0.004)
Control group mean	0.446	0.040	0.224	0.188
Observations	87,864	87,865	87,924	86,958
R^2	0.64	0.44	0.49	0.62
Panel C: Main analysis sample, excluding villages intersected by a canal				
Below canal	0.054***	0.045***	0.002	0.008
	(0.008)	(0.005)	(0.007)	(0.005)
Control group mean	0.427	0.033	0.215	0.185
Observations	55,816	55,794	55,834	55,396
R^2	0.66	0.33	0.51	0.67
Panel D: Main analysis sample, additional control for distance to canal				
Below canal	0.044***	0.048***	0.001	0.000
	(0.008)	(0.006)	(0.006)	(0.005)
Control group mean	0.428	0.032	0.213	0.189
Observations	76,618	76,622	76,678	75,888
R^2	0.62	0.40	0.47	0.64
Panel E: Main analysis sample, only long and straight canals with canal-segment fixed effects				
Below canal	0.073***	0.088***	-0.014	-0.002
	(0.018)	(0.014)	(0.017)	(0.009)
Control group mean	0.428	0.032	0.213	0.189
Observations	20,872	20,865	20,869	20,760
R^2	0.71	0.54	0.55	0.55
Panel F: Main analysis sample, no land area weighting				
Below canal	0.074***	0.109***	-0.011*	-0.017***
	(0.007)	(0.007)	(0.007)	(0.005)
Control group mean	0.428	0.032	0.213	0.189
Observations	76,618	76,622	76,678	75,888
R^2	0.63	0.35	0.46	0.49
Panel G: Main analysis sample, Conley standard errors				
Below canal	0.075***	0.099***	-0.011	-0.004
	(0.014)	(0.012)	(0.009)	(0.007)
Control group mean	0.428	0.032	0.213	0.189
Observations	76,614	76,619	76,675	75,884
R^2	0.02	0.03	0.00	0.00

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

Notes: This table demonstrates the robustness of results in Table 3 (following Equation 5.1) for irrigation outcomes. Panel A uses all settlements ≤10km and ±2.5−−50m from the nearest canal in elevation. Panel B employs the same sample definition as our main analysis sample, but defines settlement elevation using the 25th percentile. Panel C excludes settlements intersected by a canal branch, while Panel D adds an additional control for distance to the nearest canal. Panel E uses only settlements whose nearest canal segment is ≥5km (long) and ≤1.2 sinuosity (straight) and uses canal-segment rather than subdistrict fixed effects. Panel F shows our main specification without land area weights while Panel G shows our main specification but with Conley standard errors to account for spatial correlation.

Table A6: Regression discontinuity results for agricultural outcomes (robustness)

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water-intensive crops (any)	Mechanized farm equip. (share of all HHs)
Panel A: All canal-area settlements, minus donut hole					
Below canal	0.040***	0.026***	0.066***	0.029***	0.005**
	(0.005)	(0.008)	(0.011)	(0.008)	(0.002)
Control group mean	0.554	7.704	7.228	0.561	0.048
Observations	121,955	121,924	121,525	95,430	116,883
R^2	0.59	0.81	0.69	0.74	0.31
Panel B: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation					
Below canal	0.030***	0.037***	0.065***	0.022***	0.001
	(0.005)	(0.009)	(0.013)	(0.007)	(0.002)
Control group mean	0.614	7.693	7.241	0.556	0.058
Observations	95,055	94,990	94,711	75,928	91,121
R^2	0.61	0.82	0.72	0.71	0.33
Panel C: Main analysis sample, excluding villages intersected by a canal					
Below canal	0.018***	-0.001	0.041***	0.022**	0.001
	(0.005)	(0.009)	(0.011)	(0.009)	(0.002)
Control group mean	0.593	7.714	7.220	0.541	0.055
Observations	61,614	61,583	61,441	48,857	58,801
R^2	0.62	0.83	0.73	0.74	0.31
Panel D: Main analysis sample, additional control for distance to canal					
Below canal	0.018***	-0.005	0.059***	0.014	0.001
	(0.006)	(0.009)	(0.012)	(0.009)	(0.002)
Control group mean	0.595	7.692	7.210	0.555	0.057
Observations	83,512	83,450	83,190	65,691	79,972
R^2	0.61	0.83	0.71	0.72	0.31
Panel E: Main analysis sample, only long and straight canals with canal-segment fixed effects					
Below canal	0.036***	0.040**	0.072**	0.044**	-0.005
	(0.013)	(0.018)	(0.032)	(0.018)	(0.005)
Control group mean	0.595	7.692	7.210	0.555	0.057
Observations	23,189	23,177	23,080	19,416	22,181
R^2	0.68	0.85	0.82	0.79	0.36
Panel F: Main analysis sample, no land area weighting					
Below canal	0.031***	0.038***	0.023*	0.036***	0.002
	(0.004)	(0.010)	(0.013)	(0.007)	(0.002)
Control group mean	0.595	7.692	7.210	0.555	0.057
Observations	83,512	83,450	83,190	65,691	79,972
R^2	0.66	0.74	0.68	0.69	0.24
Panel G: Main analysis sample, Conley standard errors					
Below canal	0.027***	0.017	0.071***	0.027**	0.002
	(0.008)	(0.014)	(0.019)	(0.011)	(0.002)
Control group mean	0.595	7.692	7.210	0.555	0.057
Observations	83,510	83,448	83,188	65,666	79,970
R^2	0.09	0.01	0.01	0.01	0.00

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

Notes: This table demonstrates the robustness of results in Table 3 (following Equation 5.1) for agriculture outcomes. Panel A uses all settlements ≤10km and ±2.5−−50m from the nearest canal in elevation. Panel B employs the same sample definition as our main analysis sample, but defines settlement elevation using the 25th percentile. Panel C excludes settlements intersected by a canal branch, while Panel D adds an additional control for distance to the nearest canal. Panel E uses only settlements whose nearest canal segment is ≥5km (long) and ≤1.2 sinuosity (straight) and uses canal-segment rather than subdistrict fixed effects. Panel F shows our main specification without land area weights while Panel G shows our main specification but with Conley standard errors to account for spatial correlation.

Table A7: Regression discontinuity results for non-farm outcomes (robustness)

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, landless HHs)	Consumption pc (log) (log, land-owning HHs)
Panel A: All canal-area settlements, minus donut hole						
Below canal	0.190***	0.002	0.003***	-0.001	0.008	0.025***
	(0.025)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)
Control group mean	5.139	0.091	0.058	0.019	9.583	9.783
Observations	123,823	115,207	115,207	115,207	113,814	113,038
R^2	0.42	0.28	0.20	0.26	0.47	0.54
Panel B: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation						
Below canal	0.133***	0.001	0.002	0.000	0.004	0.022***
	(0.025)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)
Control group mean	5.317	0.091	0.059	0.020	9.607	9.822
Observations	96,599	90,392	90,392	90,392	88,757	88,670
R^2	0.46	0.29	0.21	0.27	0.46	0.58
Panel C: Main analysis sample, excluding villages intersected by a canal						
Below canal	0.113***	0.001	0.003*	-0.002**	0.007	0.012**
	(0.026)	(0.002)	(0.002)	(0.001)	(0.007)	(0.006)
Control group mean	5.220	0.088	0.058	0.018	9.594	9.804
Observations	62,433	57,831	57,831	57,831	56,944	56,934
R^2	0.44	0.33	0.19	0.29	0.44	0.54
Panel D: Main analysis sample, additional control for distance to canal						
Below canal	0.090***	-0.001	0.002	-0.002**	0.000	0.012**
	(0.028)	(0.003)	(0.002)	(0.001)	(0.006)	(0.006)
Control group mean	5.239	0.090	0.059	0.020	9.603	9.812
Observations	84,763	79,291	79,291	79,291	77,791	77,720
R^2	0.43	0.26	0.19	0.28	0.46	0.55
Panel E: Main analysis sample, only long and straight canals with canal-segment fixed effects						
Below canal	0.183***	-0.001	0.003	-0.002	-0.011	-0.002
	(0.056)	(0.006)	(0.004)	(0.002)	(0.015)	(0.015)
Control group mean	5.239	0.090	0.059	0.020	9.603	9.812
Observations	23,559	21,879	21,879	21,879	21,617	21,498
R^2	0.56	0.28	0.22	0.33	0.48	0.60
Panel F: Main analysis sample, no land area weighting						
Below canal	0.113***	-0.001	0.001	-0.001*	-0.003	0.020***
	(0.019)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)
Control group mean	5.239	0.090	0.059	0.020	9.603	9.812
Observations	84,763	79,291	79,291	79,291	77,791	77,720
R^2	0.37	0.17	0.12	0.19	0.34	0.47
Panel G: Main analysis sample, Conley standard errors						
Below canal	0.154***	0.001	0.003	-0.001	0.002	0.021***
	(0.034)	(0.003)	(0.002)	(0.001)	(0.006)	(0.006)
Control group mean	5.239	0.090	0.059	0.020	9.603	9.812
Observations	84,763	79,290	79,290	79,290	77,788	77,716
R^2	0.04	0.00	0.00	0.00	0.01	0.01

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

Notes: This table demonstrates the robustness of results in Table 3 (following Equation 5.1) for non-farm outcomes. Panel A uses all settlements $\leq$10km and $\pm 2.5--50$m from the nearest canal in elevation. Panel B employs the same sample definition as our main analysis sample, but defines settlement elevation using the 25th percentile. Panel C excludes settlements intersected by a canal branch, while Panel D adds an additional control for distance to the nearest canal. Panel E uses only settlements whose nearest canal segment is $\geq$5km (long) and $\leq$1.2 sinuosity (straight) and uses canal-segment rather than subdistrict fixed effects. Panel F shows our main specification without land area weights while Panel G shows our main specification but with Conley standard errors to account for spatial correlation.

Table A8: Regression discontinuity results for education outcomes (robustness)

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (literate share of pop.)
Panel A: All canal-area settlements, minus donut hole				
Below canal	0.020***	0.020***	0.015***	0.013***
	(0.003)	(0.003)	(0.002)	(0.002)
Control group mean	0.455	0.296	0.184	0.556
Observations	116,821	116,821	116,821	123,823
R^2	0.58	0.57	0.53	0.59
Panel B: Canal-area settlements balanced on ruggedness, using 25th percentile settlement elevation				
Below canal	0.015***	0.013***	0.011***	0.009***
	(0.003)	(0.003)	(0.002)	(0.002)
Control group mean	0.485	0.319	0.201	0.575
Observations	91,077	91,077	91,077	96,599
R^2	0.57	0.56	0.54	0.60
Panel C: Main analysis sample, excluding villages intersected by a canal				
Below canal	0.009**	0.009***	0.008***	0.009***
	(0.004)	(0.003)	(0.003)	(0.002)
Control group mean	0.472	0.308	0.193	0.566
Observations	58,762	58,762	58,762	62,433
R^2	0.56	0.55	0.52	0.57
Panel D: Main analysis sample, additional control for distance to canal				
Below canal	0.006	0.007**	0.006**	0.007***
	(0.004)	(0.003)	(0.002)	(0.002)
Control group mean	0.476	0.311	0.196	0.569
Observations	79,924	79,924	79,924	84,763
R^2	0.56	0.55	0.52	0.57
Panel E: Main analysis sample, only long and straight canals with canal-segment fixed effects				
Below canal	0.006	0.011	0.011*	0.016***
	(0.009)	(0.007)	(0.006)	(0.006)
Control group mean	0.476	0.311	0.196	0.569
Observations	22,170	22,170	22,170	23,559
R^2	0.58	0.58	0.56	0.56
Panel F: Main analysis sample, no land area weighting				
Below canal	0.012***	0.013***	0.010***	0.008***
	(0.003)	(0.003)	(0.002)	(0.002)
Control group mean	0.476	0.311	0.196	0.569
Observations	79,924	79,924	79,924	84,763
R^2	0.44	0.43	0.42	0.44
Panel G: Main analysis sample, Conley standard errors				
Below canal	0.013***	0.013***	0.010***	0.011***
	(0.004)	(0.003)	(0.003)	(0.003)
Control group mean	0.476	0.311	0.196	0.569
Observations	79,922	79,922	79,922	84,763
R^2	0.02	0.02	0.02	0.02

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

Notes: This table demonstrates the robustness of results in Table 3 (following Equation 5.1) for education outcomes. Panel A uses all settlements ≤10km and ±2.5−−50m from the nearest canal in elevation. Panel B employs the same sample definition as our main analysis sample, but defines settlement elevation using the 25$^{\text{th}}$ percentile. Panel C excludes settlements intersected by a canal branch, while Panel D adds an additional control for distance to the nearest canal. Panel E uses only settlements whose nearest canal segment is ≥5km (long) and ≤1.2 sinuosity (straight) and uses canal-segment rather than subdistrict fixed effects. Panel F shows our main specification without land area weights while Panel G shows our main specification but with Conley standard errors to account for spatial correlation.

Table A9: Regression discontinuity results for command area boundary specification

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)
Inside command area	0.113***	0.164***	-0.012	-0.028***
	(0.012)	(0.013)	(0.012)	(0.009)
Control group mean	0.469	0.047	0.285	0.148
Observations	43,172	43,134	43,167	42,695
R^2	0.68	0.41	0.50	0.46

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Inside command area	0.031***	0.125***	0.027	0.009	0.002
	(0.008)	(0.019)	(0.025)	(0.015)	(0.003)
Control group mean	0.657	7.569	7.336	0.653	0.050
Observations	48,190	48,245	48,139	41,594	45,860
R^2	0.66	0.77	0.75	0.72	0.32

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)
Inside command area	0.200***	0.003	0.002	0.003*	0.013*
	(0.039)	(0.004)	(0.002)	(0.002)	(0.008)
Control group mean	5.763	0.084	0.058	0.020	9.707
Observations	48,809	45,004	45,004	45,004	46,130
R^2	0.59	0.24	0.20	0.25	0.52

Panel D. Education outcomes

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (literate share of pop.)
Inside command area	0.025***	0.020***	0.016***	0.014***
	(0.006)	(0.005)	(0.004)	(0.004)
Control group mean	0.446	0.297	0.179	0.550
Observations	45,848	45,848	45,848	48,809
R^2	0.63	0.59	0.54	0.65

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

Notes: This table reports the alternative regression discontinuity estimates using the command area boundary identification strategy. This model uses distance to the nearest command area boundary as the running variable, with settlements inside the command area being able to access canal irrigation while those outside the command area cannot. Each outcome variable is estimated separately, with the β_1 coefficient of the estimate reported in the first row with stars indicating its significance and the standard error below in parentheses. The control group mean (weighted by land area), the number of observations with non-missing data for the particular outcome variable, and the adjusted R^2 for each regression estimate are also shown.

Table A10: Regression discontinuity results for primary outcomes by sinuosity

Panel A. Sinuosity ≤ 1.1

Min. canal length (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
2	0.074***	0.075**	0.113*	-0.011	-0.096	16,031
	(0.021)	(0.034)	(0.064)	(0.008)	(0.080)	
5	0.100***	0.092*	0.144*	-0.018*	-0.166*	8,691
	(0.030)	(0.050)	(0.085)	(0.010)	(0.100)	
10	0.123***	0.051	0.196	-0.012	0.015	4,043
	(0.047)	(0.090)	(0.141)	(0.017)	(0.101)	

Panel B. Sinuosity ≤ 1.2

Min. canal legnth (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
2	0.063***	0.064***	0.155***	-0.003	-0.042	35,643
	(0.014)	(0.024)	(0.043)	(0.005)	(0.079)	
5	0.073***	0.072**	0.183***	-0.001	-0.094	23,559
	(0.018)	(0.032)	(0.056)	(0.006)	(0.089)	
10	0.094***	0.050	0.224***	0.001	-0.001	14,147
	(0.026)	(0.048)	(0.081)	(0.009)	(0.115)	

Panel C. Sinuosity ≤ 1.5

Min. canal length (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
2	0.065***	0.061***	0.132***	-0.003	0.021	61,330
	(0.011)	(0.015)	(0.032)	(0.004)	(0.065)	
5	0.073***	0.072**	0.183***	-0.001	-0.094	23,559
	(0.018)	(0.032)	(0.056)	(0.006)	(0.089)	
10	0.094***	0.050	0.224***	0.001	-0.001	14,147
	(0.026)	(0.048)	(0.081)	(0.009)	(0.115)	

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

Notes: This table tests for robustness of the results from the "long and straight canal segment" sample to different parameter choices. It reports the regression discontinuity estimates following Equation 5.1 for four main outcomes and the ruggedness balance test using only settlements whose nearest canal segment is $\geq x \in 2,5,10$ km (long) and $\leq y \in 1.1,1.2,1.5$ sinuosity (straight), while also using canal-segment rather than subdistrict fixed effects. A canal segment is defined as a single line feature from the GIS data. We use long and straight canal segments for this exercise because their shape indicates that they were not constructed in such a way to include or exclude specific settlements due to political, economic, or other endogenous characteristics. Each outcome variable is estimated separately, with the β_1 coefficient of the estimate reported in the first row with stars indicating its significance and the standard error below in parentheses. The number of observations with non-missing data for the particular outcome variable is also shown.

Table A11: Regression discontinuity results for primary outcomes (sensitivity analysis)

Panel A. Regression discontinuity bandwidth

Bandwidth (m)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
25	0.066***	0.052***	0.102***	0.000	-0.022	81,765
	(0.009)	(0.012)	(0.030)	(0.003)	(0.041)	
50	0.068***	0.080***	0.154***	0.001	0.053	84,763
	(0.008)	(0.012)	(0.028)	(0.002)	(0.068)	
75	0.073***	0.073***	0.172***	0.000	0.032	84,402
	(0.008)	(0.012)	(0.029)	(0.002)	(0.073)	

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)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
10%	0.066***	0.050***	0.148***	0.000	0.010	51,044
	(0.011)	(0.016)	(0.036)	(0.003)	(0.036)	
25%	0.075***	0.071***	0.154***	0.001	0.053	84,763
	(0.008)	(0.012)	(0.028)	(0.002)	(0.068)	
50%	0.075***	0.068***	0.171***	0.001	-0.069	108,664
	(0.008)	(0.011)	(0.023)	(0.002)	(0.059)	

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)	Total emp. (share of adult pop.)	Ruggedness (TRI)	Sample size
5	0.081***	0.064***	0.193***	-0.003	0.046	55,571
	(0.012)	(0.016)	(0.034)	(0.004)	(0.053)	
10	0.075***	0.071***	0.154***	0.001	0.053	84,763
	(0.008)	(0.012)	(0.028)	(0.002)	(0.068)	
15	0.069***	0.078***	0.164***	0.000	0.015	101,436
	(0.007)	(0.012)	(0.038)	(0.002)	(0.050)	

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

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

Table A12: Comparison to distant settlements for irrigation outcomes (robustness)

	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, 0-10km above-canal settlements, no outliers dropped				
Below-canal minus	0.057***	0.086***	-0.003	-0.014*
above-canal settlements	(0.015)	(0.011)	(0.007)	(0.008)
Above-canal minus	0.012*	0.004	0.004	0.004
distant settlements	(0.007)	(0.003)	(0.006)	(0.005)
Control group mean	0.465	0.071	0.216	0.185
Observations	103,844	104,060	104,034	103,279
R^2	0.60	0.18	0.39	0.76
Panel B. Entropy balance, 0-10km above-canal settlements, 5% outliers dropped				
Below-canal minus	0.054***	0.094***	-0.003	-0.024***
above-canal settlements	(0.016)	(0.014)	(0.010)	(0.008)
Above-canal minus	0.010	0.007	0.009	-0.005
distant settlements	(0.008)	(0.005)	(0.007)	(0.005)
Control group mean	0.437	0.066	0.212	0.167
Observations	55,491	55,667	55,653	55,130
R^2	0.63	0.18	0.44	0.78
Panel C. Entropy balance, 0-5km above-canal settlements, 2.5% outliers dropped				
Below-canal minus	0.050***	0.075***	-0.005	-0.011
above-canal settlements	(0.016)	(0.011)	(0.010)	(0.010)
Above-canal minus	0.017	0.007	0.012	-0.003
distant settlements	(0.011)	(0.005)	(0.009)	(0.007)
Control group mean	0.450	0.086	0.200	0.170
Observations	35,878	35,957	35,937	35,722
R^2	0.58	0.19	0.40	0.76
Panel D. Entropy balance, 0-20km above-canal settlements, 2.5% outliers dropped				
Below-canal minus	0.044***	0.086***	-0.006	-0.027***
above-canal settlements	(0.014)	(0.011)	(0.007)	(0.007)
Above-canal minus	0.024**	0.005	0.015**	0.005
distant settlements	(0.010)	(0.005)	(0.007)	(0.008)
Control group mean	0.454	0.071	0.211	0.178
Observations	59,036	59,121	59,163	58,617
R^2	0.66	0.22	0.43	0.78

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

Notes: This table reports the spillover analysis estimates following Equation 5.2 for irrigation outcomes to test the robustness of our main results presented in Table 5. Each panel reports γ_1 (below-canal) and $-\gamma_2$ (distant settlements) estimates from an independent regression. Panels A and B define above-canal settlements as 0–10km distance from the canal while varying the threshold for excluding outliers. Panels C and D define above-canal settlements as 0–5km and 0–20km from the canal respectively. Weights were calculating using entropy balancing and district fixed effects are used in all specifications.

Table A13: Comparison to distant settlements for agricultural outcomes (robustness)

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of all HHs)
Panel A. Entropy balance, 0-10km above-canal settlements, no outliers dropped					
Below-canal minus	0.017***	-0.003	0.059***	0.030*	0.002
above-canal settlements	(0.006)	(0.013)	(0.020)	(0.016)	(0.002)
Above-canal minus	-0.003	-0.007	-0.024	0.025	-0.002
distant settlements	(0.007)	(0.012)	(0.016)	(0.016)	(0.002)
Control group mean	0.574	7.755	7.331	0.680	0.043
Observations	114,967	115,148	114,879	88,268	110,421
R^2	0.54	0.84	0.59	0.66	0.30
Panel B. Entropy balance, 0-10km above-canal settlements, 5% outliers dropped					
Below-canal minus	0.024***	0.026**	0.033	-0.011	0.002
above-canal settlements	(0.009)	(0.012)	(0.024)	(0.014)	(0.003)
Above-canal minus	0.001	0.004	-0.030	0.030	0.003
distant settlements	(0.008)	(0.010)	(0.021)	(0.020)	(0.002)
Control group mean	0.574	7.857	7.358	0.633	0.036
Observations	61,979	61,968	61,857	45,927	58,985
R^2	0.55	0.89	0.60	0.73	0.31
Panel C. Entropy balance, 0-5km above-canal settlements, 2.5% outliers dropped					
Below-canal minus	0.018**	0.012	0.089***	0.021	0.004
above-canal settlements	(0.008)	(0.013)	(0.019)	(0.016)	(0.003)
Above-canal minus	-0.012	0.018	-0.066**	0.077***	0.000
distant settlements	(0.010)	(0.026)	(0.031)	(0.026)	(0.003)
Control group mean	0.544	7.818	7.332	0.629	0.042
Observations	40,977	40,955	40,830	30,508	39,107
R^2	0.55	0.88	0.59	0.71	0.34
Panel D. Entropy balance, 0-20km above-canal settlements, 2.5% outliers dropped					
Below-canal minus	0.018***	0.014	0.058**	0.027*	0.000
above-canal settlements	(0.007)	(0.013)	(0.026)	(0.014)	(0.003)
Above-canal minus	0.007	0.011	-0.033	0.033	0.004*
distant settlements	(0.008)	(0.014)	(0.025)	(0.025)	(0.002)
Control group mean	0.567	7.805	7.309	0.667	0.039
Observations	66,683	66,641	66,450	50,242	63,589
R^2	0.59	0.87	0.55	0.66	0.35

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

Notes: This table reports the spillover analysis estimates following Equation 5.2 for agricultural outcomes to test the robustness of our main results presented in Table 5. Each panel reports γ_1 (below-canal) and $-\gamma_2$ (distant settlements) estimates from an independent regression. Panels A and B define above-canal settlements as 0–10km distance from the canal while varying the threshold for excluding outliers. Panels C and D define above-canal settlements as 0–5km and 0–20km from the canal respectively. Weights were calculating using entropy balancing and district fixed effects are used in all specifications.

Table A14: Comparison to distant settlements for non-farm outcomes (robustness)

	Population density (log)	Total emp (share of adult pop.)	Services emp (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption pc (log, all HHs)
Panel A. Entropy balance, 0-10km above-canal settlements, no outliers dropped					
Below-canal minus	0.191***	0.002	0.003**	0.000	0.021***
above-canal settlements	(0.024)	(0.002)	(0.001)	(0.001)	(0.005)
Above-canal minus	0.034	-0.001	0.000	-0.001	-0.005
distant settlements	(0.029)	(0.002)	(0.001)	(0.001)	(0.008)
Control group mean	5.665	0.088	0.057	0.021	9.653
Observations	116,773	107,081	107,081	107,081	111,140
R^2	0.32	0.12	0.10	0.18	0.41
Panel B. Entropy balance, 0-10km above-canal settlements, 5% outliers dropped					
Below-canal minus	0.181***	0.001	0.002	0.001	0.024***
above-canal settlements	(0.034)	(0.002)	(0.001)	(0.001)	(0.007)
Above-canal minus	0.056*	-0.002	-0.001	-0.001	0.003
distant settlements	(0.029)	(0.004)	(0.001)	(0.002)	(0.009)
Control group mean	5.604	0.080	0.051	0.020	9.625
Observations	62,712	57,153	57,153	57,153	59,287
R^2	0.26	0.16	0.08	0.26	0.38
Panel C. Entropy balance, 0-5km above-canal settlements, 2.5% outliers dropped					
Below-canal minus	0.161***	0.001	0.003*	0.000	0.029***
above-canal settlements	(0.032)	(0.003)	(0.002)	(0.001)	(0.009)
Above-canal minus	0.042	-0.007	-0.001	-0.002	0.009
distant settlements	(0.042)	(0.006)	(0.002)	(0.003)	(0.013)
Control group mean	5.515	0.094	0.055	0.023	9.640
Observations	41,450	38,045	38,045	38,045	39,321
R^2	0.24	0.18	0.09	0.25	0.41
Panel D. Entropy balance, 0-20km above-canal settlements, 2.5% outliers dropped					
Below-canal minus	0.157***	0.002	0.002	0.001	0.022***
above-canal settlements	(0.029)	(0.003)	(0.002)	(0.001)	(0.007)
Above-canal minus	0.041	0.001	-0.001	0.001	0.001
distant settlements	(0.031)	(0.003)	(0.002)	(0.002)	(0.008)
Control group mean	5.620	0.079	0.052	0.019	9.646
Observations	67,473	62,127	62,127	62,127	63,954
R^2	0.29	0.15	0.10	0.22	0.44

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

Notes: This table reports the spillover analysis estimates following Equation 5.2 for non-farm outcomes to test the robustness of our main results presented in Table 5. Each panel reports γ_1 (below-canal) and $-\gamma_2$ (distant settlements) estimates from an independent regression. Panels A and B define above-canal settlements as 0–10km distance from the canal while varying the threshold for excluding outliers. Panels C and D define above-canal settlements as 0–5km and 0–20km from the canal respectively. Weights were calculating using entropy balancing and district fixed effects are used in all specifications.

Table A15: Comparison to distant settlements for land ownership outcomes (robustness)

	Consumption pc (log, landless HHs)	Consumption pc (log) (log, land-owning HHs)	Middle school ed. (share of landless pop.)	Middle school ed. (share of land-owning pop.)
Panel A. Entropy balance, 0-10km above-canal settlements, no outliers dropped				
Below-canal minus	0.005	0.021***	0.012***	0.025***
above-canal settlements	(0.005)	(0.006)	(0.003)	(0.005)
Above-canal minus	-0.012*	-0.004	0.000	0.004
distant settlements	(0.007)	(0.009)	(0.004)	(0.005)
Control group mean	9.523	9.752	0.258	0.363
Observations	107,339	106,445	107,114	106,826
R^2	0.37	0.40	0.43	0.52
Panel B. Entropy balance, 0-10km above-canal settlements, 5% outliers dropped				
Below-canal minus	0.006	0.024***	0.015***	0.031***
above-canal settlements	(0.005)	(0.006)	(0.003)	(0.005)
Above-canal minus	-0.012	0.007	0.002	0.006
distant settlements	(0.009)	(0.010)	(0.005)	(0.006)
Control group mean	9.487	9.730	0.250	0.353
Observations	56,837	57,490	56,689	57,611
R^2	0.35	0.41	0.45	0.53
Panel C. Entropy balance, 0-5km above-canal settlements, 2.5% outliers dropped				
Below-canal minus	0.006	0.026***	0.012***	0.027***
above-canal settlements	(0.007)	(0.010)	(0.003)	(0.005)
Above-canal minus	-0.001	0.025*	0.000	0.006
distant settlements	(0.014)	(0.015)	(0.007)	(0.009)
Control group mean	9.506	9.739	0.252	0.349
Observations	37,872	38,061	37,782	38,151
R^2	0.36	0.40	0.42	0.54
Panel D. Entropy balance, 0-20km above-canal settlements, 2.5% outliers dropped				
Below-canal minus	0.012*	0.025***	0.015***	0.026***
above-canal settlements	(0.007)	(0.008)	(0.004)	(0.005)
Above-canal minus	-0.009	0.005	-0.001	0.005
distant settlements	(0.009)	(0.010)	(0.005)	(0.006)
Control group mean	9.514	9.742	0.259	0.361
Observations	61,456	61,808	61,320	61,979
R^2	0.38	0.43	0.42	0.53

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

Notes: This table reports the spillover analysis estimates following Equation 5.2 for outcomes disaggregated by land ownership to test the robustness of our main results presented in Table 5. Each panel reports γ_1 (below-canal) and $-\gamma_2$ (distant settlements) estimates from an independent regression. Panels A and B define above-canal settlements as 0–10km distance from the canal while varying the threshold for excluding outliers. Panels C and D define above-canal settlements as 0–5km and 0–20km from the canal respectively. Weights were calculating using entropy balancing and district fixed effects are used in all specifications.

Table A16: Effect of canals on town size and population (robustness)

	Population (log)		Town Existence (pop. 5,000)	
*Panel A. Add State * Year Fixed Effects*				
Command area in town catchment area	0.077***		0.031***	
(binary treatment)	(0.028)		(0.013)	
Share of town catchment area in command area		0.234***		0.091***
(continuous treatment)		(0.041)		(0.018)
Observations	302691	64260	302691	64260
R^2		0.840		0.720
Panel B. Drop Years After 1990				
Command area in town catchment area	0.096***		0.025*	
(binary treatment)	(0.038)		(0.016)	
Share of town catchment area in command area		0.248***		0.070***
(continuous treatment)		(0.056)		(0.024)
Observations	231436	52080	231436	52080
R^2		0.830		0.700
Panel C. Define Catchment Area as 10 km Radius				
Command area in town catchment area	0.101***		0.029**	
(binary treatment)	(0.032)		(0.014)	
Share of town catchment area in command area		0.250***		0.083***
(continuous treatment)		(0.038)		(0.017)
Observations	301519	49464	301519	49464
R^2		0.830		0.700
Panel D. Define Catchment Area as 30 km Radius				
Command area in town catchment area	0.107***		0.028**	
(binary treatment)	(0.030)		(0.014)	
Share of town catchment area in command area		0.289***		0.076***
(continuous treatment)		(0.047)		(0.018)
Observations	301966	74244	301966	74244
R^2		0.830		0.700

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

Notes: The table shows results from alternate specifications of Equation 5.3. The setup is identical to Table 6, with the following changes: Panel A adds state * year fixed effects to the estimation. Panel B drops locations where the first canal has a completion year later than 1990 (where canal dates are more likely to refer to rehabilitation than to initial completion). Panel C defines the treatment based on the amount of canal coverage within 10 km of each town (rather than 20 km as in Table 6). Panel D does the same, with a 30 km radius.

Table A17: Impact of district-level canal expansion on in-migration

Panel A. Migration by period of canal expansion

	Outcome: is a migrant			
Treatment period	1941-1981	1951-1981	1961-1981	1991-2021
Canal coverage gain	0.101***	0.066**	0.070**	0.027
	(0.031)	0.027)	(0.033)	(0.024)
Base year canal coverage	0.048***	0.060**	0.061**	0.030
	(0.017)	0.016)	(0.015)	(0.018)
Control group mean	0.241	0.241	0.241	0.241
Observations	624,628	624,628	624,628	624,628
R^2	0.02	0.02	0.02	0.02

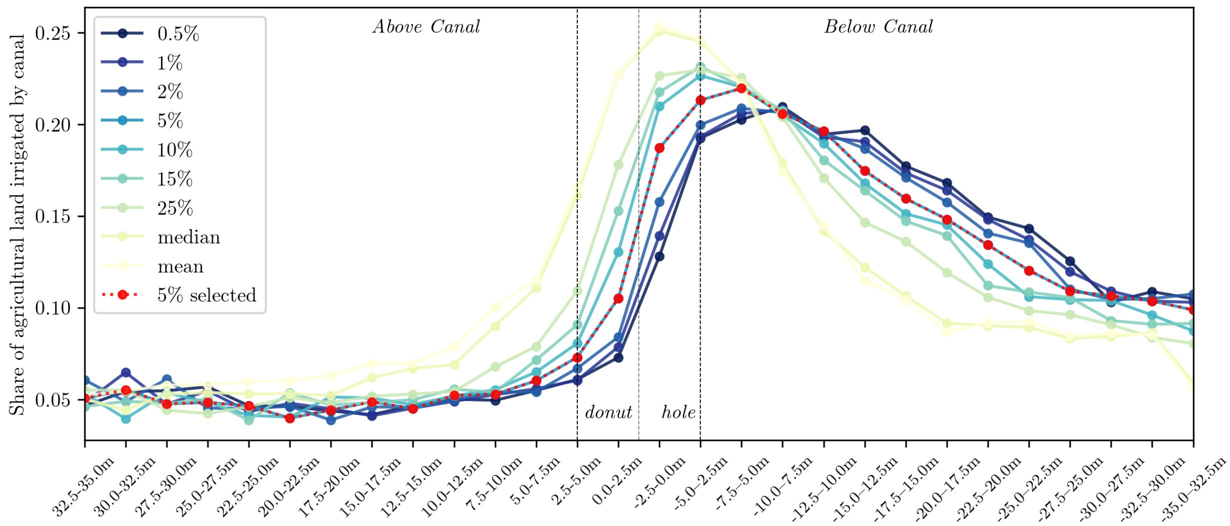
Panel B. Migration by origin and destination

	Outcome: Is a migrant from a rural area		Outcome: Is a migrant from an urban area	
Sample	Rural	Urban	Rural	Urban
Canal coverage gain (1951-1981)	0.066***	0.120***	0.002	-0.023
	(0.023)	0.044)	(0.009)	(0.025)
Base year canal coverage (1951)	0.040***	0.076***	0.013	-0.003
	(0.015)	0.028)	(0.005)	(0.016)
Control group mean	0.191	0.166	0.020	0.121
Observations	419,677	206,380	419,677	206,380
R^2	0.02	0.01	0.01	0.01

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

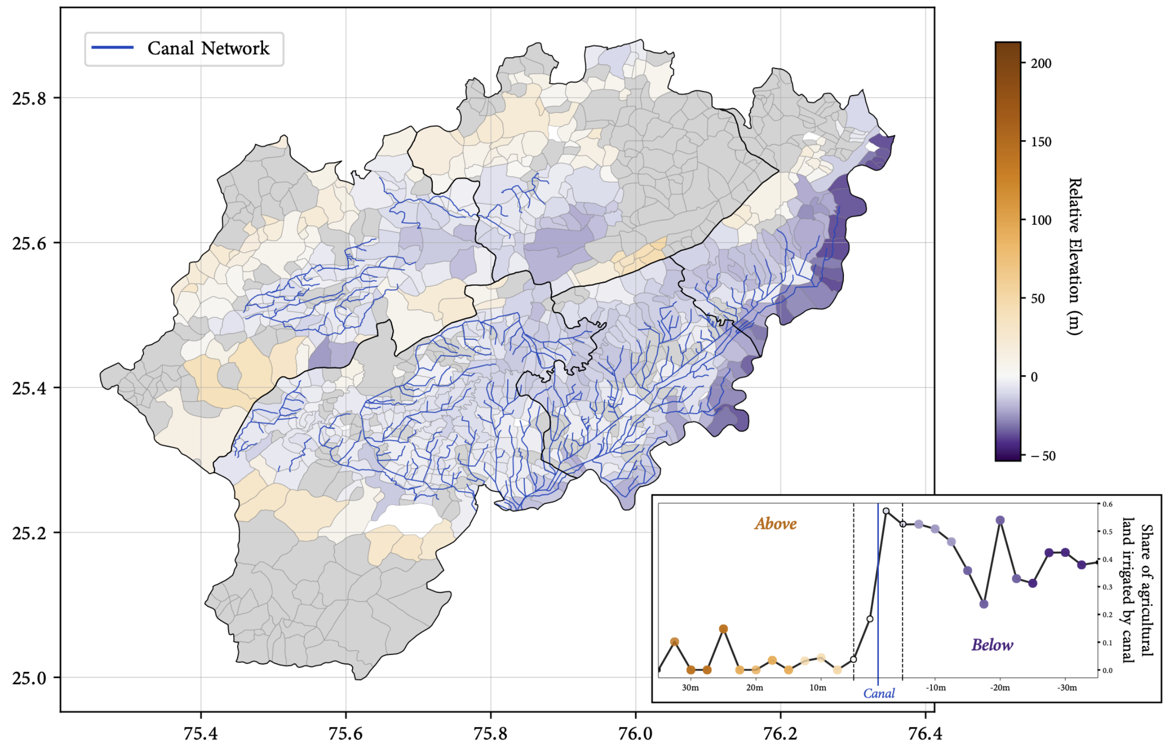
Notes: This table shows the estimated effect of district-level canal coverage gain on in-migration, using data from the 1987–88 (43rd) round of India's National Sample Survey. The estimating equation $y_i = \alpha_0 + \alpha_1 canal\_gain_i + \alpha_2 canal\_baseline_i + \zeta_i + \epsilon_i$, where y_i is the outcome of interest, the treatment variable $canal\_gain_i$ measures the share of the area of district i that gained coverage by canal command areas over the period of interest, $canal\_baseline_i$ controls for the share of the area of district i that had canal coverage at the start of the period of interest, and ζ_i is a state fixed effect. Panel A defines the outcome variable as a binary variable for whether the respondent has migrated to their place of residence. The first three columns consider different periods of extensive canal construction (1941–81, 1951–81, and 1961–81), all ending before the survey was conducted in 1987–88. The fourth column is a placebo exercise that tests for whether canals built in 1991–2021, after the data were collected, has any "effect" on the outcome. In Panel B, we test for the source of migration. We use the period 1951–81 and define the outcome as a binary for being a migrant from a rural area (first two columns) or being a migrant from an urban area (second two columns), estimated separately for respondents living in rural areas (columns 1 and 3) and in urban areas (columns 2 and 4). Standard errors are clustered at the district level.

Figure A1: Calculating the relative elevation of each settlement



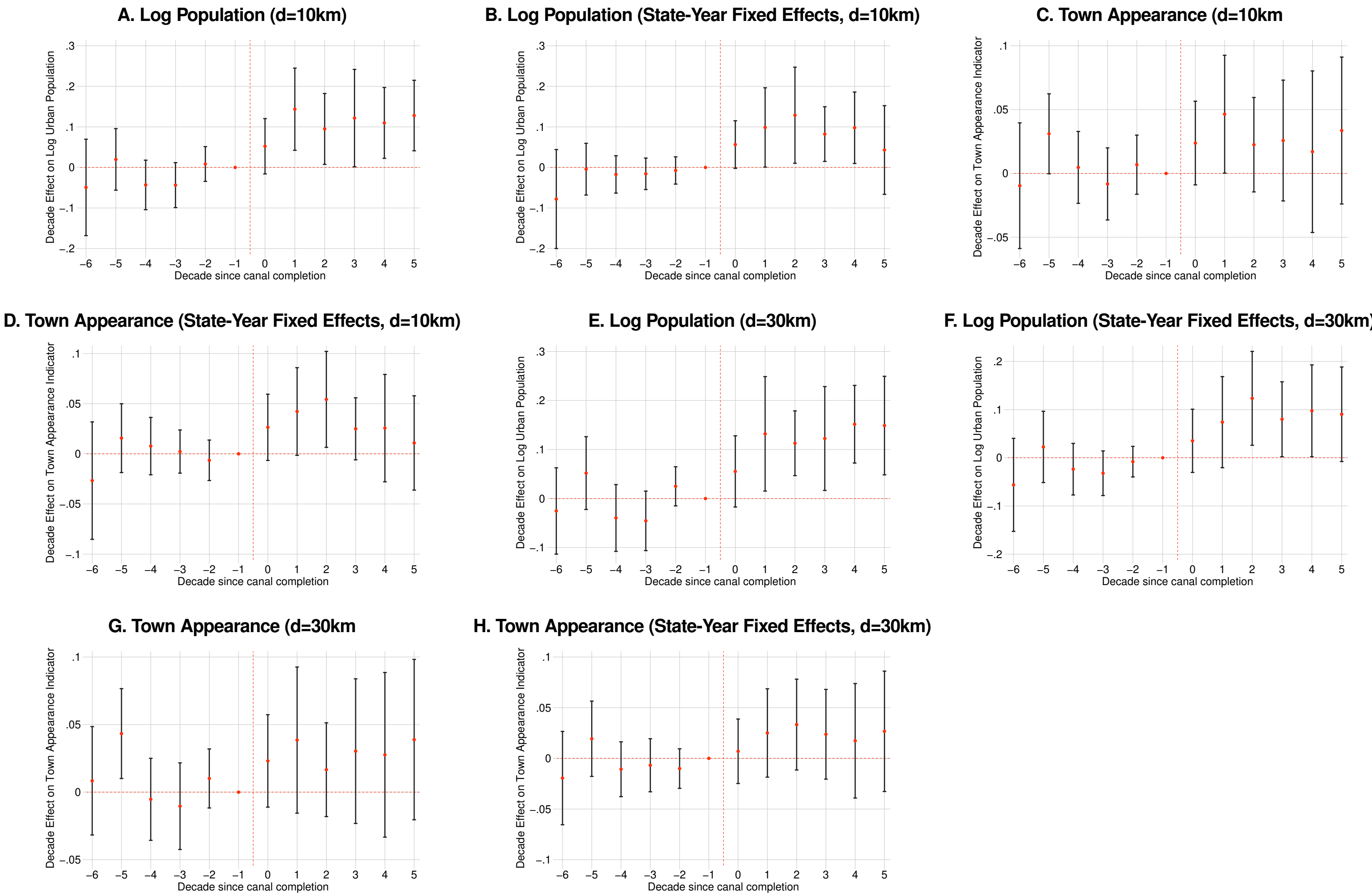
Notes: Each line in this figure uses a different moment of the distribution of elevation in a settlement polygon to define the relative elevation between that settlement and the nearest canal. The elevation of the nearest canals is parameterized by the elevation of the single closest point. Share of 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 5$^{\text{th}}$ percentile to define settlement elevation in our preferred specification.

Figure A2: Relative elevation RDD empirical strategy



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

Figure A3: Effects of canal construction on town appearance and size (alternate distance thresholds)



Notes: The figure shows difference-in-differences plots (calculated following De Chaisemartin and d'Haultfoeuille (2020)) describing the effect of canal construction on urban population (Panels A, B, E, and F) and town emergence (Panels C, D, G, and H). The estimation is identical to Figure 5, but defining the zone in which canals can influence towns as a 10km (Panels A-D) or 30km (Panels E-H) radius circle around the town, instead of 20km as in Figure 5. All estimations include town and decade fixed effects and standard errors are clustered at the district level.