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

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

September 16, 2021

Abstract

How and when do improvements in agricultural productivity translate into development and structural transformation? An extensive literature has addressed these questions, dating back to the earliest days of development economics. More recently, a new literature has focused on generating well-identified empirical evidence in settings where high-quality micro data allows for the study of local shocks to productivity. Most of this literature has explicitly or implicitly assumed that labor movements across space are small, such that local productivity shocks drive within-location movements of labor across sectors. This may be a reasonable restriction for short-run analyses, but long-run outcomes may be very different due to labor mobility. In this paper, we draw on rich data that allow us to examine the long-run effects of agricultural productivity gains on structural transformation, in an environment where spatial mobility of labor is a potentially important margin of impact. Specifically, we examine irrigation canals constructed over the last 150 years in India. In the long run, canal areas have substantially higher land productivity and population density than nearby non-canal areas, consistent with higher population growth through some combination of natural increase and net in-migration. However, we see no change in the share of the workforce outside of agriculture (or even in agroprocessing). Consumption gains have accrued only to landowners, though the landless have made some education gains. These observations do not necessarily imply that agricultural productivity increases have little impact on structural transformation. Indeed, we show that structural transformation instead occurs through higher growth rates in nearby towns. Our findings suggest that in the long run, agricultural growth may drive structural transformation through these spatially diffuse processes. We should not necessarily expect to see transformation occurring within the most agriculturally productive rural areas; instead, labor mobility may produce a new spatial equilibrium in which labor moves across space rather than across sectors within a location.

*We are grateful for persistent, patient, and creative research assistance from Sam Besse, Aditi Bhowmick, Kritarth Jha, Toby Lunt, Shraddha Mandi, and Sankalp Sharma. Gayatri Acharya, Richard Damania, Anju Gaur, Ijsbrand de Jong, Stuti Sharma, Vivek Srivastava and Esha Zaveri at the World Bank provided invaluable insight into the water and agriculture sectors in India, as well as assistance in obtaining data. Financial assistance from the World Bank Agriculture Global Practice and Research Support Budget, as well as Emergent Ventures (Mercatus Center, George Mason University), is gratefully acknowledged.

†Johns Hopkins SAIS, sasher@jhu.edu

‡Development Data Lab, campion@devdatalab.org

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

¶Dartmouth College, paul.novosad@dartmouth.edu

1 Introduction

The link between agricultural productivity and structural transformation has long been a central concern of development economics (Lewis, 1954; Johnston and Mellor, 1961; Mellor, 1986; Schultz, 1964). Authors such as Johnston and Mellor (1961), echoed later by Mellor (1986) and Timmer (1988), argued that agricultural productivity growth was an essential precursor for broader structural transformation. This early literature held that productivity growth in agriculture could have the seemingly paradoxical effect of shrinking the agricultural sector as a share of the total economy. Building on the insight that food is an essential good for the poor, agricultural development economists invoked a class of models in which countries that are unproductive in agriculture must, precisely as a result, devote large shares of labour and other resources to meet their food needs. T.W. Schultz (1953) referred to this phenomenon as the "food problem". The same mechanism lies at the heart of more recent work such as Gollin et al. (2007) and Gollin et al. (2002), which relied on non-homothetic preferences as the main driver of structural transformation. But the link between agricultural productivity growth and structural change is not simply a product of models with non-homotheticity; it also emerges in other models where productivity growth leads to endogenous changes in the relative price of agricultural goods. For instance, in the model of Ngai and Pissarides (2004), differential productivity growth across sectors leads to structural change, with the type of change depending on the extent of substitutability between sectors.

However, a conflicting strand of literature has long challenged the conceptual (lower) basis of these agriculture-driven models of structural change. One critique has been that these models depend heavily on a closed economy assumption. Implicitly, these models assume that the relative prices of agricultural and non-agricultural goods are determined endogenously. The key mechanism driving structural change in Gollin et al. (2007) or Alvarez-Cuadrado and Poschke (2011) is a change in the relative price of agriculture. But as noted by Matsuyama (1992), relative prices for a tradeable good are exogenous to a small open economy. To the extent that agricultural goods are tradeable, we should expect to see increases in productivity pulling resources into agriculture in an open economy when sectoral productivity rises. In Matsuyama (1992), there is a single input (labor), so increases in agricultural productivity pull workers into agriculture – precisely the opposite of the mechanism proposed by Johnston and Mellor (1961). The same intuition applies in the somewhat more complicated framework of Bustos et al. (2016), with more than one input into production. In this framework, agricultural productivity growth will tend to lead to an expansion of the sector, but whether this increases or reduces the use of labor in agriculture will depend on the factor bias of technical change.

What these two strands of literature both assume, however, is that the responses to agricultural productivity growth will primarily take place within economies that have fixed endowments of factors. Agricultural productivity growth induces price changes and the reallocation of existing factors. What these models lack is any response to the supply of factors. Economies may be open to trade in goods, but in this literature there is typically no provision for the movement of factors. Neither labor nor capital flows into economies that experience changes in agricultural productivity.

In this paper, we relax the assumption of limited factor mobility. We consider a setting in which agricultural productivity growth may drive transformation through the movement of people – not just across sectors within an economy, but across locations. Specifically, we consider a setting where spatially distinct rural economies (which we term "municipalities") are open to human mobility – as is the case in rural India, where labor movements across locations are commonplace and quantitatively large, at least in the long run. In this environment, labor mobility may be a primary channel of adjustment to agricultural productivity growth.

Our work differs from earlier research that assumes low labor mobility, primarily because we consider much longer time periods. Where much of the existing work implicitly or explicitly assumes that agricultural productivity shocks do not induce any movements of people across locations (e.g., (Foster and Rosenzweig, 1996; Foster and Rosenzweig, 2004b; Foster and Rosenzweig, 2004a; Bustos et al., 2016; Emerick, 2018; Moscona, 2020)), we find that labor mobility offers an important margin of adjustment. The analysis we report here focuses on one of the most quantitatively significant episodes of agricultural productivity change of the past two centuries. Specifically, we study the long-run effects of investments in the universe of India's major and medium irrigation canals. This network of irrigation canals – essentially artificial rivers and streams that carry water into dryland areas – span thousands of miles and have transformed agricultural productivity in many regions of the country. The canal network has been especially important in enabling irrigation during the relatively dry winter cropping season. Canals were historically the most important source of irrigation in India, and even in the 21st century, they are the second largest source of irrigation in India after groundwater. The network of canals that we study in this paper provides irrigation to agricultural communities with over a quarter billion inhabitants. According to our calculations, in 2011 fully 52% of rural Indians lived within 10 km of a major or medium canal. Over half of the canals that we study were built before the 1980s and some are more than 100 years old, making them ideal for studying long-run responses to major gains in agricultural productivity.

These canals provide an ideal natural experiment for our question because they drive sustained and large differences in agricultural productivity across otherwise similar locations. We show that, in the long run, the substantial productivity effects of canals are equilibrated almost entirely through the movement of labor. In addition to higher land productivity and agricultural expansion on the extensive margin, rural canal-treated areas experience substantial growth in population density. However, there are no overall changes in consumption. Structural transformation takes place entirely through the formation and growth of towns at some distance from canal areas; we estimate a tight zero on the rural non-agricultural labor share. These results are consistent with a model where labor mobility is high (in the long run), but towns have some productivity advantage for non-farm work, such as reduced transport costs to wider markets or agglomeration externalities.

Our analysis proceeds in three parts. First, we obtain causal identification on the direct effects of irrigation from the physical nature of canals — their placement is determined by engineering specifications and topography, and water from the canals then flows downhill so that it treats only lower-elevation locations. Locations a short distance away from a canal but a few meters of elevation higher than the canal will thus experience no irrigation benefit and will serve as a control group for the irrigation treatment.[1] We exploit this sharp difference in irrigation access in a regression discontinuity design, where elevation relative to the canal is the running variable.

We find a sharp effect of canal irrigation on agricultural outcomes right at the canal elevation

[1]In principle, these "control" areas might benefit from recharged groundwater, but we find no evidence of increased pump irrigation in the control areas.

boundary. Locations with canal access report greater agricultural land area, with a greater share of irrigated land and more water-intensive crops planted. A satellite-derived measure of crop yields suggests that land productivity is substantially higher as well. The yield effects are observed entirely in the relatively dry winter (Rabi) season: canals improve water access to a second cropping season but generate no significant differences during the summer (Kharif) growing seasons, when monsoon rains provide sufficient water.

When we look at socioeconomic outcomes, canal areas have substantially higher population density than non-canal areas, but no significant difference in the share of workers employed across the agriculture, manufacturing, and services sectors; nor do we find significant differences in two-digit or three-digit sectors such as agroprocessing. Canal areas have higher living standards than neighboring non-canal areas, but this difference is driven entirely by land-owning households. The standardized effect size of the canal on population density swamps all of the other effect sizes, suggesting that the productivity effects of canals are primarily equilibrated through labor flows across space, rather than sectoral reallocation within locations. The main effect of canals is to draw in additional labor to treated locations, with little effect on the magnitude of the nonfarm sector. These results are robust to a range of alternate specifications, including an alternate regression discontinuity using distance to the officially designated command area boundary of the canal.[2]

We next consider spillovers, which are central to our analysis. Above-canal locations (the control group above) are likely to be economically integrated with canal locations due to their proximity; they may trade goods and services with canal areas, and migration may occur across the treatment boundary, especially in the long run. Given this economic integration, it would be possible for canals to have large local effects—on wages or non-farm work, for example—without generating any differential impacts between locations just above and just below the canal. To measure canal treatment spillovers, we compare both the treated and untreated groups above to a third set of municipalities

[2] The command area is the engineers' definition of the total area that theoretically has access to irrigation water from a given canal. This definition exploits finer details of local topography, but risks endogeneity if command area boundaries were drawn such that they intentionally include or exclude certain locations. The elevation definition performs better in terms of balance, but we find similar overall results under either definition. In both definitions, we exclude locations directly on the treatment border where the spillovers may be greatest and identification of treatment status is most subject to error.

that are further from the canal but are otherwise similar in terms of climate, topography, and crop suitability.[3] We call these "distant" locations. Differences between above-canal control locations and distant locations would be suggestive of direct spillovers across the canal treatment threshold.

The spillover analysis provides support for the identification strategy: distant locations have similar sources of irrigation, agricultural productivity, and population density to above-canal locations, suggesting that irrigation access does not spill over the canal boundary, nor does migration. There are also no spillovers in terms of non-farm work or consumption among the landless; the null effect of the non-agricultural labor share in the RDD is sustained in the spillover analysis as well. However, landowners in above-canal areas have higher consumption than landowners in distant areas. Once spillovers are taken into account, landowners in above-canal areas experience spillover consumption gains that are about one third of the estimated treatment effect in below-canal areas.

These results suggest that labor supply is elastic across a wide geographic range. If workers could move only short distances such that all of the new workers in below-canal areas came from nearby above-canal areas, we might see wages rise in both of these areas relative to distant locations. The absence of spillover effects suggests that, in the long run, rural India draws upon a much wider labor market. All this said, the comparison to distant locations is not as well-identified as the differences between above- and below-canal municipalities, and there could be other omitted variables that are obscuring large spillover effects.[4]

Finally, we use historical urban population data to examine changes in urbanization, exploiting the construction dates of canals in a difference-in-differences setup. In locations close to command areas, urban population growth rates rise substantially and towns are more likely to emerge in the decades following canal construction. Structural transformation has thus taken place in the vicinity of the canal investments, but the growth of non-agricultural economies has taken place in a geographically concentrated manner, with little effect on the large share of the economy that has remained rural.

[3] These municipalities are between 15 and 50 km from the nearest canal; we find similar results with alternate distance choices, suggesting that municipalities 15–50 km from the canal are plausibly not directly affected by the canal.

[4] For instance, suppose that spillover effects to above-canal municipalities are large, but canal-proximate areas suffer from some other disadvantage compared with distant municipalities, and these two effects cancel each other out. However, we do not find large differences in observables between above-canal and distant areas.

We interpret our results in the context of a model where non-agricultural production can occur anywhere, but towns have a productivity advantage from agglomeration. In the absence of labor mobility, Hicks-neutral technical change in agriculture would affect local wages and through them, the size of the non-agricultural sector.[5] This model is a close analog to that used in much of the prior literature on structural transformation. When we allow for labor mobility across space, the increased agricultural demand for labor can be met by migrants from other areas, rather than drawing workers out of manufacturing. Labor mobility thus dampens the link between agricultural productivity and rural manufacturing. If the pool of available labor from other areas is large, even a small agglomeration advantage in urban areas is sufficient to attract nearly all non-agricultural firms. Our results suggest that, in the long run, labor mobility has indeed been the primary mechanism of adjustment to the increased land productivity associated with canals.

The key contribution of this paper is to study the causal relationship between agricultural productivity change and structural transformation, with a focus on changes in the long run. Our findings suggest that the labor mobility channel, which appears small in shorter run studies, may be dominant in the long run. We also show that the geography of structural transformation is important: the industrialization effects of agricultural changes arise via growth of urban areas rather than through a shift of rural workers into rural manufacturing.

This paper extends a substantial literature linking technical change in agriculture to industrialization. Foster and Rosenzweig (1996) and Foster and Rosenzweig (2004b) studied the impact of the Green Revolution on structural change and educational attainment in a panel of villages across India, finding that agricultural gains raised wages and inhibited industrialization. Bustos et al. (2016) found similar effects of the introduction of a second harvesting season for maize in Brazil, but opposite effects in soybeans where the agricultural productivity change was labor-augmenting rather than land-augmenting. We conduct a similar analysis of a major land-augmenting productivity improvement

[5] Bustos et al. (2016) show that land-augmenting and Hicks-neutral technical change in agriculture increase competition for workers from manufacturing in such a model, while labor-augmenting technical change can have the opposite effect. Canals are best thought of as land-augmenting technical change since they effectively make land available for agriculture for a longer season. We use the Hicks-neutral assumption as it is consistent with the prior literature and it generates the same outcomes.

in agriculture, but over a much longer time horizon that permits greater net movement of labor, and we find much smaller effects on the rural economy. We also find that structural transformation tends to occur outside the village, and occurs even in response to land-augmenting technical change.

It is notable that much of the prior empirical work has aimed to study competition for workers between the farm and non-farm sectors in models that shut down the labor mobility channel. This is partly for the reason that mobility is typically much lower in the short- to medium-run periods examined in prior studies. Indeed, in an extension of their main results, Bustos et al. (2016) find that about one-third of the shift out of agricultural employment in soybean areas occurred via migration, over only a 10-year sample period. Our much longer run analysis suggests that it can be the primary adjustment channel to agricultural change. Indeed, the very nature of structural transformation around the world has involved moving billions of people from farms to cities, sometimes across large distances.[6]

Our results are also consistent with studies finding that the barriers to rural industrialization are high. Asher and Novosad (2020) and Burlig and Preonas (2021) find that major investments in rural roads and electrification respectively have limited effects on non-farm activity in India.[7] Infrastructure investments in villages may improve well-being and may motivate in- and out-migration, but are unlikely to cause substantial changes in manufacturing opportunities in those villages. Our results are also consistent with long run evidence that the Green Revolution had substantial positive effects on structural change (Gollin et al., 2021); our analysis suggests that this process may have been driven by the growth of cities, rather than in the rural areas directly experiencing agricultural productivity change.

Our work also adds to a growing literature estimating the impacts of agricultural innovation, including canal irrigation. Jones et al. (2020) study canal irrigation in Rwanda using an elevation-based RD, with a focus on understanding why households do not completely substitute inputs to canal-irrigated land. In a concurrent paper, Blakeslee et al. (2021) study canals in India using a command area

[6] While the idea that permanent migration is rare in India is widespread in the literature, this claim appears to arise from focusing on the set of rural men who migrate for work. But over 25% of women have changed residence at least once in their lives, and migration rates for men approach 15% (Kone et al., 2018). Since we only observe population density in the present, we cannot distinguish migration from other mechanisms of population change. However, we show that there are at least no contemporary effects on fertility either in below- or above-canal areas.

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

distance RD similar to our elevation RD specification.[8] While they find similar reduced form effects on population density and (lack of) structural change in canal-irrigated villages, their analysis does not consider spillovers and is focused primarily on evaluating canals as infrastructure investments, rather than as drivers of long-run agricultural change.[9]

Finally, our results echo several findings in other work studying structural transformation. Bustos et al. (2020) find that land rents were invested in cities, consistent with our finding that agriculture-driven structural change may largely take place away from the farms directly experiencing productivity shocks. Our results also echo the predictions of Foster and Rosenzweig (2004a), which suggested that agricultural productivity shocks have substantially different effects on landowners and the landless, consistent with our findings.[10]

Our paper proceeds as follows. Section 2 provides background on India's irrigation canals. Section 3 describes our conceptual framework for understanding how agricultural productivity gains spill over into other economic outcomes. Section 5 presents our multiple empirical strategies. Section 4 describes the data sources. Section 6 presents our results. Section 7 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 irrigation through canals of various types. It was estimated at that the end of the 19th century, India's 12 million hectares of irrigated land amounted to 4 times that of the United States and 6 times that of Egypt (Shah, 2011). The British oversaw the construction of vast canal networks, often privately funded and yielding high returns until the end of the Raj in 1947. Canals were used

[8] This is analogous to the secondary identification strategy we use to show robustness of the RD.

[9] Blakeslee et al. (2021) differ from us in arguing that canals *reduce* city size directly in their command area. Our analysis uses a time series of urban populations and finds that new towns emerge in the vicinity of canals; we do not examine direct effects of irrigated areas on towns, given that some towns may only exist as a result of the canal.

[10] In the Foster and Rosenzweig (2004a) model, productivity gains in agriculture can raise or lower schooling for landless individuals. They can lower landless schooling because children of landed households may increase schooling and exit the low-skill labor market, raising demand for unskilled labor and drawing landless children out of school. However, increased demand for schooling among the landed can increase the supply of schools, with positive spillover effects on the landless. Our results are consistent with the second case, in that we find schooling benefits to the landless, albeit smaller than those to the landed.

to divert water from India's major rivers to its arid regions, where they facilitated settlement of otherwise uninhabitable land, such as with the Punjab Canal Colonies (Douie, 1914).

After gaining independence, the Government of India prioritized canals as it sought to avoid mass hunger during a period of high population growth (Mukherji, 2016). Later, they were a key part of the strategy to provide sufficient irrigation for the input-intensive high yield variety crops that powered India's Green Revolution.

While groundwater eclipsed canals as India's preeminent source of irrigation in the 1970s, surface irrigation remains critical to the livelihoods of millions of farmers across India. In recognition of the importance of canals, the central government launched the Accelerated Irrigation Benefit Program (AIBP) in 1997. By 2011 it had spent more than $7.5 billion to help finish stalled large-scale irrigation projects (Shah, 2011). More recently, states such as Madhya Pradesh have dramatically increased canal irrigation (at least according to official statistics) through increasing maintenance, distributing water from tail to head to reduce head-tail disparities, investments in last mile distribution networks, reducing political interference, and building cooperation with farmer organizations (Shah and Steinberg, 2016). According to the most recent estimates, canals account for approximately 1/4 of the net irrigated area in India (Jain et al., 2019), although estimates vary according to the methodology.

We study the long-run impacts of the universe of India's major and medium canals. Figure 1 plots the length of such canals completed in each five year bin since 1850. While many canals were built in the 19th and early 20th century, construction takes off following Indian independence in 1947, although the post-independence canals are generally shorter than those constructed under the British Raj. By 2012, the main year in which we measure outcomes, 52% of India's 600,000 villages were within 10 km of a major or medium irrigation canal, with a median canal construction start year of 1968 and completion year of 1981 – this paper thus studies the impacts of canals that are generally several decades old.

3 Conceptual Framework

Our paper focuses on India's rural economy. Our conceptual framework is informed by both quantitative data and extensive qualitative experience. A stylized view is that India's rural economy consists of a large number of predominantly local sub-economies that are embedded in a larger national economy.

Each of the local economies features an expanse of agricultural land, divided into villages, typically with a larger market town that serves as an economic center. The villages that surround each market town are mostly small, with an average size of approximately 1000 inhabitants, most of whom work in agriculture. Agricultural land is in general privately owned and managed. Most farms are small (as discussed by Foster and Rosenzweig (2017)), and many land owners work on their own land. Farms may also hire workers from a large pool of landless workers, who comprise the majority of the labor force.

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

Consumer preferences are defined over the agricultural and non-agricultural goods. As is standard, we assume a non-homotheticity that implies poor consumers will spend larger shares of their income on the agricultural good than richer consumers. Preferences over the non-agricultural good distinguish between goods produced in different locations. Consumers view non-agricultural goods produced in their own village as closely substitutable with the goods produced in the nearby town. Non-agricultural goods from the rest of the economy, however, are only weakly substitutable with local goods. This assumption reflects the fact that the non-agricultural goods include some that are less than fully tradable (e.g., retail services).

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

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

We are interested in the impacts of irrigation canals both in the short run and the long run. As other researchers have argued, labor mobility in rural India faces significant frictions in the short run (Foster and Rosenzweig, 2007; Munshi and Rosenzweig, 2016) although Foster and Rosenzweig (2007) find male out-migration from villages of over 20% when considering a longer (17 year) period. Our conceptual framework thus considers three distinct time periods: a baseline before the construction of the canals; a short run after canals have been constructed but before labor has adjusted to the changes in productivity; and a long run in which a spatial equilibrium holds, and in which real wages are equalized across locations. This long-run labor mobility is a key feature of our theoretical framework and also of our empirical work. Allowing for labor mobility leads to important differences relative to the models of Matsuyama (1992) or Bustos et al. (2016).

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

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

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

What determines the location of structural change? In other words, what determines where we see a take-off in non-agricultural production activity? In our spatial framework, each local economy consists of a set of villages that are connected, in a hub-and-spoke arrangement, to a central town. Trade between villages and towns faces a transport cost friction. Towns are connected to one another (and to the external Indian economy) frictionlessly. This setup captures the reality that movement between villages is costly, and towns function as central meeting and marketing places. Within the local economy, this means that a sale of goods from one village to another would accrue a double transport cost, relative to a sale between village and town. The model implies that increasing demand for non-agricultural goods within a region will be met by some combination of production in villages and in their central town, with the precise allocation depending on the relative productivity levels, the substitutability of town goods and village goods, and the transport costs between towns and villages. When town productivity is high and transport costs are low, and town goods are highly substitutable with village goods, we would expect to see increases in demand met through production in the towns. At the other extreme, when town productivity is only slightly higher than village productivity, and transport costs are high, increases in demand might be met through production in villages. The model permits both possibilities; it is an empirical question for this paper to see how increases in demand for non-agricultural goods may lead to structural change in different locations.

4 Data

To estimate the varied impacts of canal irrigation on current local economic outcomes, we assemble recent high resolution data on the universe of firms, households, and municipalities (villages and towns) in India. The backbone of our village/town-level dataset contains economic and population data from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG). The SHRUG creates a village/town-level identifier that is unique across time and space, allowing the merging of disparate datasets (Asher et al., 2021) at a very high spatial resolution. Using the SHRUG, we have combined various national censuses, GIS data on canals, satellite imagery, and a range of other data sources.

The 2011 Population Census contains many of our irrigation, agricultural, and demographic outcome variables. Village-level land use data allow us to calculate total cultivated land as a share of total village area and total, tubewell and canal irrigated areas, each as a share of the village's total cultivated area. The Population Census also contains the top three crops grown in each village. If cotton, sugarcane, rice, or paddy are in that list, the village is considered to grow a water-intensive crop.[11] Since municipalities may vary in their physical area, our preferred measure for population is population density (inhabitants per km^2).[12]

The 2012 Socioeconomic and Caste Census (SECC), for which we have microdata covering every household in India, gives us a range of economic outcomes at the individual level, which we aggregate to the municipality level. We predict consumption per capita from household-level asset and earnings data using the small area estimation methodology of Elbers et al. (2003). We also calculate the share of the adult population that has achieved at least primary, middle, and secondary education.[13]

[11] These agricultural variables are available in the Census Village Directory but not Town Directory, and thus all analysis using them is restricted to rural areas.

[12] Population density is calculated using the total municipality population divided by the area of the municipality GIS polygon shape (area in km^2) as opposed to the noisier area reported in the Population Census.

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

Because the SECC is microdata, we are able to calculate these outcomes by household characteristics, such as land ownership.

The 2013 Economic Census is a complete enumeration of all nonfarm economic establishments in India, which we use to calculate non-agricultural activity for each municipality. To calculate employment as a share of the adult population, we use the total population reported in the Population Census multiplied by the share of the population over the age of 18.[14] We use the National Industrial Classification codes of firms in the Economic Census to calculate the share of the adult population specifically employed in the manufacturing, service, and agroprocessing sectors.

In the absence of directly-measured municipality-level agricultural productivity data, we use the Enhanced Vegetation Index (EVI), a satellite-derived measure of biomass often used as a proxy for agricultural productivity. We calculate productivity for both the monsoon (Kharif), late May through early October, and winter (Rabi) season, late December through late March. For each season, we define the productivity value as the mean of the first six weeks of the season subtracted from the maximum value reached in the entire season. Taking this difference ensures that we subtract out the signal from any background vegetation that is non-agricultural and does not contribute to agricultural productivity. We then take the mean for the years 2011-13 (corresponding to our other outcome datasets), and log transform to address outliers and ease interpretation of effects.

All spatial data on canals and their command areas comes from the Ministry of Water Resources. Canal construction dates were obtained from the Management Information System of Water Resources Projects of the Central Water Commission in India, the India Water Resources Information System (WRIS)[15], and manually researched canal construction dates. These dates were merged with the spatial canal data using canal project code or by canal name.

Using municipality polygon GIS data, we extract the distribution of elevation contained in each municipality from Shuttle Radar Topography Mission (SRTM) raster data. Following Riley et al. (1999) and Nunn and Puga (2012), we calculate the ruggedness of a location's topography using

[14] 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 and then subtract that from the total to get the adult population.

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

the Terrain Ruggedness Index (TRI); TRI measures ruggedness as the average square difference in elevation between a pixel and its eight surrounding pixels, which we average across all pixels in a municipality. Using these same data, we compute the distance from every municipality centroid to the nearest canal, command area, river, town, and coast.

For more details on the construction and validation of the predicted consumption and agricultural productivity proxy variables, please see Asher and Novosad (2020) and Asher et al. (2021).

5 Empirical Strategy

Testing for the long-run impacts of increasing agricultural productivity is challenging because of the endogeneity of agricultural productivity differences. Large, costly infrastructural investments such as canals have generally been targeted to areas with political favor and high returns to irrigation. We are able to estimate the causal effects of such infrastructure due to the high spatial resolution nature of our data, using regression discontinuity designs to compare outcomes in nearby areas that differ only in that treatment areas have access to canal irrigation and control areas do not. We employ a secondary empirical strategy that compares canal areas to more distant areas to assess the extent of local spillovers, such as in labor markets, from canal villages to proximate areas. Finally, we use a difference-in-differences design with panel data on town population to test for the effects of canals on urban growth.

5.1 Regression Discontinuity

Our primary strategy exploits the gravity-driven nature of canals. Because water delivery from canal to field physically depends on gravity, land must be below the canal in terms of elevation to be able to access water from that canal. Municipalities that are just above the canal will not be able to access the water.

To measure the relative elevation between a municipality and canal, we calculate elevation measures for both. For the canal, we select the elevation on the nearest point of a canal to a given municipality. The municipality is characterized by a distribution of elevation values from pixels inside its polygon. We select the 5th percentile of the distribution as the point value to define elevation, which strongly predicts the difference in canal irrigation between treatment and control areas (see Figure A2 for a comparison of different potential definitions of elevation).

Equation 5.1 shows the regression discontinuity design (RDD) specification using the relative elevation strategy. y_i is the variable of interest in municipality i and T_i is the dummy variable with a value of 1 for municipalities below the elevation of the canal. $Elev_i^{below}$ is the difference in elevation between a treated municipality and the nearest canal while $Elev_i^{above}$ is the same value for non-treated municipalities. $Geog_i$ is a control for geographic fundamentals (ruggedness and mean rainfall) and $\alpha_{subdistrict}$ is the subdistrict fixed effect to ensure villages are only being compared within similar geographic regions. We cluster standard errors at the subdistrict level to account for spatial correlation.

$$y_i = \beta_1 T_i + \beta_2 Elev_i^{below} + \beta_3 Elev_i^{above} + \beta_4 Geog_i + \alpha_{subdistrict} + \epsilon_i \tag{5.1}$$

For robustness, we estimate the impacts of canals using a secondary regression discontinuity design. We exploit the boundary of the command area, which is the official geographic region served by a given canal. If a municipality centroid falls within the command area, it is defined as receiving canal access while those that fall outside do not. This strategy exploits the variation in the xy-plane as opposed to the z-axis as in the relative elevation strategy. We expect locations just inside and outside a command area boundary to be similar, except for the canal access. This strategy is less clean than our primary strategy as boundaries are drawn endogenously by government officials, who may seek to favor certain types of municipalities over others.

$$y_i = \beta_1 T_i + \beta_2 Dist_i^{inside} + \beta_3 Dist_i^{outside} + \beta_4 Geog_i + \alpha_{segment} + \epsilon \tag{5.2}$$

Equation 5.2 shows the regression discontinuity specification using the command area strategy. y_i is the variable of interest and T_i is the treatment dummy variable with a value of 1 for municipalities inside the command area. $Dist_i^{inside}$ is the distance between the command area boundary and the treatment municipalities that lie inside the command area, while $Dist_i^{outside}$ is the distance between untreated municipalities outside the command area and the boundary. $Geog_i$ is a control for geographic fundamentals (ruggedness and rainfall). Command area boundaries are divided into fixed, 10km-length segments to which municipalities are matched. $\alpha_{segment}$ is this 10km segment fixed effect, ensuring that municipalities are compared over similar geographic regions. We cluster standard errors at the command-area border-segment level to account for spatial correlation.

There are 539,646 municipalities (villages and towns) in our dataset that contain non-missing data for the majority of the outcome variables. Table 1 reports summary statistics for the total sample size and average value of each outcome variable for the full dataset and each sample we have defined for our analysis. Our full sample retains only canals built before 2012, meaning we exclude projects that are reported as completed after 2012 or incomplete in the WRIS database at the time of data collection in 2019. Because these ongoing canal construction projects may already be delivering water before completion, we drop observations that would have been treated by canals completed after 2012 from the sample. For the relative elevation strategy, the full sample contains municipalities within 10km (xy-plane) and within 50m (z-axis) in elevation of the nearest canal. For the command area strategy, the full sample contains municipalities within 25km of the command are boundary. For both, we only include fixed effect groups (subdistricts or 10km-segment groups) that contain at least one village in the treatment group and one village in the control group.

The regression discontinuity design relies on positioning each village along a single dimension, either elevation or distance, using a single point to characterize its position. In reality, in our full sample, the median municipality spans 2.1km^2 in area and 18m in elevation, meaning there will inevitably be noise at the boundary: municipalities we classify as treatment will have some portion of their area too high to be reached by canals, and control municipalities may have some portion of their area below the canal. We impose a donut hole restriction to prevent the downward bias to our treatment effects that would come from misclassifying treatment status close to the cutoff. In the donut hole sample, we drop municipalities within 2.5m of the regression discontinuity cutoff for the relative elevation strategy and municipalities closer than 2.5km from the boundary for the command area strategy.

Finally, we impose a balance restriction on ruggedness. Ruggedness is our best measure of the natural, geologic features underlying variability in expected agricultural productivity that could drive different economic outcomes in canal areas even in the absence of canal irrigation. We create a balanced sample that requires a maximum percent difference in average ruggedness between the treatment and control municipalities in a fixed effect group. For the relative elevation strategy, we keep subdistricts that have less than a 25% difference in average ruggedness between their treatment and control municipalities while for the command area strategy we keep border segment groups with a less than 25% difference. This balanced sample, shown in Table 1, is our preferred sample. We test the sensitivity of our results to each of these sample selection criteria as shown in Appendix Table A6.

The majority of canal infrastructure in India was built in the 19th or early-mid 20th centuries, so we do not have consistent, detailed economic or agricultural data on municipalities prior to canal construction. We thus establish balance in our preferred sample for our preferred (relative elevation) RDD by testing for discontinuities in natural features of municipalities, which could not be influenced by canal irrigation, by estimating Equation 5.1 (except without the geographic fundamentals *Geog* term). Table 2 presents these results, showing that there are no significant agricultural advantages in canal treatment municipalities; in fact, for two of the crop suitability measures, canal treatment municipalities appear to have slightly less natural agricultural potential than control municipalities.

5.2 Testing for spillovers

While the regression discontinuity designs generate discontinuities in access to canal water, it is possible that control municipalities could, by virtue of their proximity to treatment municipalities, experience spillover effects either via increased access to irrigation (e.g. through groundwater recharge) or via market linkages (e.g. increases in labor demand). To test for these effects, we create an alternative control sample of distant municipalities within each subdistrict that lie at least 15km from the nearest canal, meaning these municipalities lie 5km farther from the nearest canal than any treatment or control municipalities in the RDD sample. We compare both treatment and control municipalities from our RDD sample to these "distant" municipalities, as described in equation 5.3:

$$y_i = \beta T + \beta C + V_i + \alpha_{subdistrict} + \epsilon_i, \tag{5.3}$$

where T is a dummy variable for municipalities in the treatment sample and C is a dummy variable for municipalities in the RDD control sample, and thus both are estimated relative to the "distant" municipalities sample. For the analysis treatment and control groups, we limit the sample to those municipalities which lie 0-10m from the nearest canal in order to focus the comparison on the municipalities driving our RDD results. The vector of municipality-level geophysical controls V_i contains average annual rainfall, distance to the nearest river, sugarcane suitability (as modeled by GAEZ), elevation, ruggedness, and square terms for elevation and ruggedness to ensure that we are comparing similar municipalities that vary primarily in their proximity to canals. We cluster standard errors at the subdistrict-level to allow for spatial correlation.

5.3 Town growth through time

The empirical strategies described in the preceding subsections estimate the effects of canals on rural areas relative to proximate and more distant rural areas, but do not incorporate an alternative channel by which canals may drive structural transformation: the growth of urban areas in the proximity of rural areas that experience gains in agricultural productivity. Because we expect such effects to be regional, the RDDs that produce large differences in agricultural productivity across short radii are not appropriate. To test for impacts on urbanization, we exploit variation in canal construction dates and examine growth and emergence of towns in their vicinity. The 2011 Population Census lists for every town the population of that town, if it was considered urban in that year, for each census dating back to 1901. Equation 5.4 describes a standard difference-in-difference model to test whether town growth and emergence are affected by canal construction in the surrounding area:

$$y_{i,t} = \beta_1 T + \zeta_i + \nu_t + \epsilon_{i,t}, \tag{5.4}$$

where $y_{i,t}$ is either an indicator for town existence, or else a measure of town population, ζ_i is a town fixed effect and ν_t is a decade fixed effect. We observe towns once each decade. When $y_{i,t}$ represents population, we define it as the log of 2000 plus the town population— this defines non-existent towns just below the average population at which towns first appear in the data. Of the 7,526 towns present in 2011, only 1,502 existed in 1911.

For each town we define a catchment area that is a ring with outer radius r and inner radius $r-10$ around the town. We define canal treatment T as the percentage of that catchment area that is overlapped by a command area. We also use an alternate specification that defines T as a binary treatment variable where $T==1$ if *any* of the town catchment area is covered by a command area. We report results for various catchment area radii r between 10 and 100km. Standard errors are clustered at the district level.

6 Results

6.1 The Effects of Canals on Treated Municipalities: Regression Discontinuity

We begin by showing how municipalities below canals benefit from significantly more irrigation than those above. For these municipalities, 6.9 percentage points more of the land under cultivation is irrigated (Table 3, Panel A), which corresponds to a 17% gain over control municipalities. Increased irrigated area is driven entirely by canals, with treatment municipalities gaining 9.2 percentage points in canal irrigated area, an almost 300% increase compared to control municipalities. No other type of irrigation contributes to the increased irrigation in municipalities below canals, as municipalities below canals have no significant difference in tubewell irrigation and or other types of irrigation.[16]

Panel B in Table 3 shows how canal irrigation brings gains to agricultural outcomes. Municipalities below canals cultivate 2.8 percentage points more of their total village area, a 5% increase over control municipalities. The benefits of these increases in irrigation and cultivated area are concentrated in the dry season, with treatment municipalities experiencing 7.4% higher Rabi agricultural productivity than control municipalities. As expected given that canals primarily deliver water during the dry season, treatment municipalities show no significant increase in monsoon season (Kharif) productivity; this is further evidence that our treatment is orthogonal to village agricultural potential. Treatment municipalities are also more likely to grow water-intensive crops and show

[16]This may reflect competing effects of canals on groundwater irrigation: on the one hand they are thought to often recharge water tables that supply tubewells (Shah, 2011), but on the other canal irrigation may reduce the demand for alternate forms of irrigation. Alternately, they recharge the groundwater table but there is no discontinuous access to groundwater at the elevation of the canal. In Section 6.3 we find that there is no meaningful difference in tubewell irrigation between above-canal (control) municipalities and villages more distant from canals, which implies that canals may not be having large effects on groundwater irrigation through water table recharge, at least in our sample.

higher rates of complementary investments in agriculture, as measured by increased household ownership of mechanized farm equipment.

Taken together, the preceding results imply that canals have major if unsurprising effects on agricultural inputs, practices, and productivity. But do these gains translate into an improved standard of living or growth in non-agricultural sectors? In Panel C we present estimates of the impacts of canals on population density, nonfarm employment, and predicted consumption. We find that the only significant effect at the 5% level is on population: by 2011, treatment municipalities have 13.1% more people per square kilometer than control municipalities. Yet we estimate a tight null effect on living standards: we can reject a 1.5% increase and -0.5% decrease in predicted consumption in canal treatment villages with 95% confidence. This may reflect symmetric countervailing forces: there is less less agricultural land per person (as the growth in land under cultivation is much smaller than the increase in population), but the land is more productive due to greater access to irrigation. Testing for impacts on structural transformation, we find no significant effect on the share of the adult population working in non-agricultural firms generally, nor when we examine employment broken out into services and manufacturing. Even agroprocessing, which we can precisely identify using National Industrial Classification codes (e.g. 103: Processing and preserving of fruits and vegetables), shows no signs of employment growth in canal villages.

Access to canal irrigation does, however, bring increases in human capital, as shown in Panel D of Table 3. Treatment municipalities show small but precise increases of approximately 1 percentage point in the share of the adult population that has completed primary, middle, and secondary school. Consistent with these outcomes derived from the Socioeconomic and Caste Census, treatment municipalities also show a 1.0 percentage point increase in the literacy rate, as measured by the Population Census. These results add to a growing body of evidence showing that increases in labor demand can drive up educational attainment, in contexts as diverse as rural call centers, and rural roads (Jensen, 2012; Heath and Mobarak, 2015; Adukia et al., 2020).

Figure 3 summarizes the RDD coefficients reported in Table 3, normalized by the standard deviation of each variable in the control sample. Blue points show the positive, significant normalized treatment effects while red points show negative, significant effects. Gray points are not significant at the 95% confidence level. Regression discontinuity results are also shown in Figure 2, with outcome variables plotted in elevation bins after residualizing on fixed effects and controls. These figures offer alternate visualizations of the results in Table 3: there are large discontinuities in irrigation measures, agricultural outcomes, population, and education at the canal elevation cutoff, but no significant effects on nonfarm employment or living standards.

Guided by our theory, we test for distributional impacts by estimating treatment effects separately for households with and without landholdings (Table 4, Figure 4). As discussed in Section 3, in the long run a spatial equilibrium may hold in which returns to mobile factors such as labor would equalize across treatment and control municipalities, while returns to fixed factors such as land would remain permanently higher in locations receiving canal water. In the absence of high resolution data on land rents and wages, we test for the returns to these factors by testing for the heterogeneity of canal treatment effects on predicted consumption by household landholding. We first show a decline in the share of the population that are land owners in treatment municipalities compared to control municipalities. Average landholdings are lower in treatment municipalities when including all households but there is no effect when condition on owning land. Taken together with our earlier results, these findings imply that the landholding population has increased by a small amount proportional to the increase in agricultural land driven by canal access, but the majority of the population increase is driven by landless households. Turning to consumption, we find that while there is no significant effect on consumption for landless households, owners of land have predicted consumption 1.9% higher in treatment municipalities (4 Panel A), significant at the 1% level and significantly different than the estimate for landless household consumption. The bottom quintile of landowners, those owning <1.32 hectares of land, also see no gains in consumption relative to their peers in nearby municipalities, and we observe the largest predicted consumption gains among households in the highest quintiles of the landholding distribution (4 Panel B).[17] Landowners also experience larger gains in education attainment, with the landless receiving about one third of the

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

benefit in the likelihood of achieving primary, middle, or secondary school education (4 Panel C).

6.2 Robustness

We test for the robustness of our main results in three ways. First, we estimate the effects of canals using the alternative command area boundary RDD described in Section 5.1, according to Equation 5.2. While we consider this a secondary identification strategy due to the potentially endogenous placement on command area boundaries, it is instructive to test whether our results hold up to a different source of variation. In Panel F of Tables A2, A3, A4, and A5, we present the results of this estimation. Even more than in the elevation RDD, canal treatment villages in the command area RDD have a much higher share of agricultural land under irrigation than in the control villages (12 percentage points, or a 22% increase on a base share of 53%). Effects on agricultural outcomes are very close in magnitude to the elevation RDD, although in the command area specification there is also a large effect during the monsoon (Kharif) growing season. While is possible that this reflects the endogeneity of command area boundaries, as canals are generally not delivering water during the monsoon season, it is also possible that the local average treatment effect (LATE) for the command area estimation is different from the relative elevation sample due to different characteristics of the complier sample. For economic outcomes, the command area strategy again yields very similar results to the elevation strategy: large effects on population density but small and insignificant impacts on nonfarm employment and predicted consumption. Finally, for education, both the elevation and command area strategies yield significant if small effects (1-2 percentage points) on the share of the population literate and completing primary, middle, and secondary school.

Second, we estimate results for the full and donut-hole samples, for both the elevation and command area specifications, as described above. Even though the former sample is contaminated by municipalities close to the regression discontinuity cutoff whose treatment status we are less confident in, and both samples by municipalities in subdistricts where there is meaningful imbalance in ruggedness between above- and below-canal municipalities, the estimated results are broadly consistent with the effects found in our preferred samples (Tables A2, A3, A4, and A5). Treatment is always associated with significantly more irrigation, land under cultivation, Rabi productivity, population density, and educational attainment. In all elevation RDD specifications, treatment municipalities grow more water-intensive crops. The null effects on non-farm employment is confirmed in most samples. All samples produce treatment effects on consumption of under 2%, with significance in some.

Finally, we test for sensitivity of the main outcomes to parameter choice in the preferred elevation RDD (Table A6). In Panel A, we find that the estimated canal treatment effects on irrigation, Rabi productivity, population density, and literacy are highly stable across bandwidths from 20m to 75m in elevation relative to the nearest canal, and that all coefficients remain significant at the 1% level regardless of bandwidth choice. In Panel B, we find that the estimated effects are also stable across sample restrictions on maximum distance from canal to municipality: reducing the maximum distance from 10 to 5 km has little effect on the coefficients, nor does increasing the distance to 15km.

6.3 Comparison to more distant municipalities

Our regression discontinuity results cleanly identify the effects of canal irrigation, but only in the context when both treatment and control municipalities are close to canals and to each other. It is possible that control municipalities experience spillovers via market linkages from canal treatment municipalities, thus biasing the estimated impacts of agricultural productivity gains. In Table 5, we present the results of Equation 5.3 where we compare both RDD treatment and control municipalities to a sample of similar but more distant municipalities, controlling for various natural village characteristics.

Considering the coefficients for the treatment sample first, we confirm the main results from the RDD. Treatment municipalities have much more irrigation (driven by canal irrigation), higher land under cultivation, and higher Rabi agricultural productivity.[18] Treatment municipalities have greater population density but no more nonfarm jobs per capita. Consumption is higher, with gains much larger for households with land.

The coefficients for the RDD sample control locations suggest that there may be some local spillovers from canal treatment to control municipalities. Compared with the sample of more distant locations, RDD sample control municipalities are not more irrigated or agriculturally productive (Rabi productiv-

[18]Kharif productivity is actually lower in treatment municipalities than in the alternate control sample, consistent either with the possibility that canals target less fertile areas or that greater Rabi irrigation induces some substitution away from Kharif cultivation.

ity is actually slightly lower, although with marginal significance). However, population densities and predicted consumption are both somewhat higher than in the distant municipalities. These differences between canal control and distant municipalities are much smaller than the differences between canal treatment and distant (approximately one fourth the size in the case of population density and one half in the case of predicted consumption). We interpret these results through our model. In the absence of agricultural spillovers to control municipalities (Panel A), the small increases in control municipality population and living standards are likely the product of local market or marriage linkages to treatment municipalities. That landless consumption is slightly (1 percentage point) higher in canal control and treatment municipalities than distant municipalities (less than 1/3 the estimated difference of 3.4 percentage points between land owners in canal treatment and distant areas) suggests that while in the long run landless labor flows across space nearly equalize living standards, spatial frictions prevent them from entirely erasing differences in consumption even many decades after canal construction.

Figures 5 and 6 show these results alongside the RDD results. The figures compare the RDD results (Columns 1 and 4) to the difference between the canal treatment and control effects calculated by Equation 5.3 (Columns 2 and 5), both measuring the difference between treated municipalities and their nearby control municipalities. Columns 3 and 6 estimate the local spillover effect by estimating the differences in outcomes between canal control and distant municipalities. That the estimates in Columns 2 and 5 closely track those in the RDD estimates in Columns 1 and 4 suggest that this exercise of comparing canal-proximate to canal-distant areas (controlling for natural features) is likely picking up the causal effect of canals on both canal treatment municipalities (through the increase in agricultural productivity) and canal control municipalities (through market and marriage linkages to canal treatment areas).

6.4 The Effects of Canals on Town Growth: Diff-in-Diffs

The increased agricultural productivity experienced by canal-irrigated municipalities does not appear to drive local economic growth in the non-agricultural sector. However, it is possible that structural transformation is induced by agricultural gains in the minority of municipalities that are urban enough to provide the conditions for nonfarm growth. In this section, we test this hypothesis in panel data on town populations going back to 1901. Bringing in data on the dates of canal construction completion, we examine how town growth changes when a canal is built nearby. Figure 7 takes the set of 2011 town locations that can be observed for 30 years before and after a canal is built (i.e. towns within 30 km of a canal that is built between 1951 and 1971), and plots their mean log population. There is a clear trend break around the date of canal construction, indicating that towns grew much more quickly in the 30 years after nearby canals were built than in the 30 years before.

Our formal estimation uses the difference-in-differences specification described by Equation 5.4, with results shown in Table 6. Columns 1 through 4 show that towns are more likely to appear in the Population Census in years following canal construction, controlling for the average rate of town appearance. This result is robust to defining town appearance as a population greater than either 5,000 or 10,000, and to both a binary canal measure (any command area within a 50 km radius of a town), or a continuous canal measure (the share of a town's circular catchment that is in a canal command area). Columns 5 and 6 demonstrate a significant intensive margin effect of canals on town log population. The continuous measure suggests that the saturation of an area with canal irrigation causes a statistically and economically significant 20% increase in the area's urban population.

How does distance affect the impact of canals on urban growth? Figure 8 plots the diff-in-diff estimate for various distance rings around town centroids. Across a range of different measures of urbanization, canals within 10km of towns have the largest effects on urban population growth. Effects rapidly attenuate with distance from canal to town, but even canals 100km from towns appear to increase their growth, implying that the increased demand for non-agricultural goods is met by distant urban production.[19] Given that even our maximum distance shows significant effects, it is likely that we are underestimating the impact of canals on city growth, as firms in cities across the country (and perhaps even world) may be meeting some of the demand. But the fact that the largest effects are estimated within 10km says that spatial frictions such as transport costs may limit the distance at which agricultural productivity translates into structural transformation.

[19] It is possible that there are other channels not captured by our parsimonious model by which agricultural productivity growth may contribute to urbanization, such as financing investments migration (Clemens, 2014) or non-agricultural firms (Bustos et al., 2020).

7 Conclusion

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

In the long run, we find that spatial equilibrium is restored primarily through substantial changes in the size of the laborer population. Decades after the canals were built, there are few differences in living standards between landless workers in canal and non-canal municipalities. However, structural transformation has taken place, with towns emerging disproportionately near canal-irrigated municipalities. Canal-fed agriculture may have competed for labor from the secondary and tertiary sectors, but that competition was outweighed by other channels.

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

Many shorter term studies have found that rising agricultural wages can deter or delay industrialization. Our study suggests that, in the long run, these effects may be tempered by changes in the labor supply. Naturally, it is difficult to compare different contexts in different times and places. Most of India's canals were built during the License Raj period, where manufacturing investments were slow and state-inhibited and may have had difficulties responding to changes in labor demand, potentially enhancing the role of mobile labor. Whether modern agricultural shocks will be equally mitigated by labor flows remains an important question for researchers.

Mobile workers pose challenges for applied empirical researchers by violating assumptions of population stability across treatment and control groups. Yet hundreds of millions of Indians report living in places other than those of their birth, and there are tens of millions of temporary migrants on top of those. Our study suggests that this large mobile population is a powerful economic force that can affect policy outcomes substantially.

Table 1: Summary statistics

	All India	Relative Elevation *Full Sample*	Relative Elevation *Donut Hole*	Relative Elevation *Balanced Sample*	Command Area *Full Sample*	Command Area *Donut Hole*	Command Area *Balanced Sample*
Sample Size	539,646	260,893	145,119	98,661	251,791	64,996	49,450
Percent Treatment	–	83	78	80	27	38	41
Total irrigated area (share of ag. land)	0.466	0.581	0.519	0.539	0.456	0.505	0.545
Canal irrigated area (share of ag. land)	0.132	0.172	0.135	0.130	0.084	0.114	0.117
Tubewell irrigated area (share of ag. land)	0.200	0.262	0.225	0.243	0.218	0.274	0.300
Other irrigated area (share of ag. land)	0.142	0.161	0.177	0.180	0.161	0.124	0.136
Agricultural land (share of total village area)	0.584	0.669	0.632	0.651	0.591	0.639	0.687
Kharif agricultural production, EVI-derived (log)	7.565	7.735	7.706	7.687	7.701	7.663	7.630
Rabi agricultural production, EVI-derived (log)	7.228	7.366	7.285	7.295	7.310	7.352	7.349
Any water intensive crop grown (any)	0.590	0.664	0.620	0.629	0.600	0.660	0.702
Mechanized farming equipment (share of households)	0.047	0.062	0.056	0.062	0.047	0.047	0.051
Population density (log)	5.510	5.736	5.584	5.622	5.518	5.795	5.960
Consumption (log)	9.719	9.747	9.741	9.753	9.694	9.687	9.704
Total nonfarm employment (share of adult pop)	0.091	0.081	0.083	0.083	0.082	0.079	0.072
Services employment (share of adult pop)	0.065	0.057	0.058	0.058	0.057	0.054	0.052
Manufacturing employment (share of adult pop)	0.020	0.020	0.021	0.022	0.019	0.019	0.019
Primary school ed attained (share of adult pop)	0.477	0.502	0.496	0.503	0.457	0.438	0.458
Middle school ed attained (share of adult pop)	0.322	0.341	0.333	0.335	0.304	0.294	0.309
Secondary school ed attained (share of adult pop)	0.198	0.215	0.211	0.212	0.189	0.180	0.190
Literacy rate (literate share of adult pop)	0.568	0.582	0.583	0.588	0.558	0.546	0.556

Notes: There are 539,646 villages and towns in our sample that have data for most relevant variables. The analysis sample contains all municipalities ≤10km from the nearest canal in distance. For the relative elevation specification, the analysis sample also requires municipalities to be within ±50m of the nearest canal in elevation. For the command area specification, the analysis sample requires municipalities to be within 25km of the command area boundary. Imbalanced fixed effect groups are dropped, meaning any subdistrict or 10km segment groups (for the relative elevation and command area strategies respectively) that contain no municipalities in either treatment or control were dropped. The donut hole sample drops all municipalities 0-2.5m from the canal in elevation, using the relative elevation specification, or within 2.5km of the command area boundary, using the command area specification. The balanced sample then imposes a balance criteria on ruggedness by dropping all subdistricts in which there is a ≥25% difference in average ruggedness between treatment and control municipalities for the relative elevation strategy and a ≥25% difference between treatment and control municipalities in 10km segment groups for the command area strategy.

Table 2: Balance

Panel A. Geophysical parameters

	Ruggedness	Annual rainfall avg. 2010-2014 (mm)	Distance to coast (km)
Below canal	-0.044	-3.226*	0.065
	(0.055)	(1.709)	(0.366)
Control group mean	4.733	1166.497	366.643
Observations	98,553	98,553	98,553
R^2	0.570	0.990	1.000

Panel B. Crop suitability

	Sugarcane	Wetland rice	Wheat	Maize
Below canal	-0.034	0.008	0.003	-0.028
	(0.021)	(0.015)	(0.005)	(0.021)
Control group mean	3.425	2.277	0.768	3.312
Observations	98,553	98,553	98,553	98,553
R^2	0.880	0.880	0.970	0.890

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

Notes: This table reports the regression discontinuity effect on several outcomes we expect to be balanced following Equation 5.1, except without the control on ruggedness. Crop suitability measures are taken from the Global Agro-Ecological Zones model that estimates expected conditions for agricultural production based on climate, soil, and terrain parameters. Model values for sugarcane, wetland rice, groundnuts, and maize assume that crops have gravity-fed irrigation and intermediate inputs (e.g. some fertilizer).

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	0.069***	0.092***	-0.006	-0.006
	(0.008)	(0.006)	(0.006)	(0.005)
Control group mean	0.436	0.031	0.213	0.198
Observations	89,628	89,467	89,539	88,757
R^2	0.610	0.380	0.480	0.640

Panel B. Agriculture outcomes

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of households)
Below canal	0.028***	0.014	0.074***	0.027***	0.004*
	(0.005)	(0.009)	(0.012)	(0.009)	(0.002)
Control group mean	0.602	7.689	7.210	0.561	0.056
Observations	96,403	97,089	96,792	75,349	92,357
R^2	0.610	0.820	0.700	0.720	0.310

Panel C. Economic outcomes

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption per capita (log)
Below canal	0.131***	-0.005	-0.004	-0.003*	0.005
	(0.017)	(0.007)	(0.005)	(0.002)	(0.005)
Control group mean	5.354	0.082	0.057	0.020	9.737
Observations	90,223	90,695	91,022	91,126	92,327
R^2	0.460	0.010	0.010	0.040	0.530

Panel D. Education outcomes

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (share of pop.)
Below canal	0.012***	0.012***	0.009***	0.010***
	(0.003)	(0.003)	(0.002)	(0.002)
Control group mean	0.481	0.316	0.200	0.572
Observations	92,268	92,268	92,268	97,879
R^2	0.570	0.560	0.530	0.590

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

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

Table 4: Regression discontinuity results for land owners vs. landless

Panel A. Land ownership overview

	Land owners (share of households)	Avg. size of land holdings (log hectares, all households)	Avg. size of land holdings (log hectares, land owners)	Consumption (log)	
				Land owners	*Landless*
Below canal	-0.022***	-0.042**	0.002	0.019***	0.000
	(0.005)	(0.018)	(0.013)	(0.005)	(0.006)
Control group mean	0.538	0.751	1.499	9.814	9.608
Observations	92,359	89,752	89,718	90,094	89,921
R^2	0.460	0.470	0.510	0.560	0.470

Panel B. Consumption distribution

	Consumption (log) 1^{st} quintile land holders	Consumption (log) 2^{nd} quintile land holders	Consumption (log) 3^{rd} quintile land holders	Consumption (log) 4^{th} quintile land holders	Consumption (log) 5^{th} quintile land holders
Below canal	0.000	0.014**	0.013*	0.021***	0.028***
	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
Control group mean	9.630	9.769	9.808	9.848	9.935
Observations	91,762	81,749	77,301	79,475	74,063
R^2	0.500	0.460	0.410	0.420	0.390

Panel C. Education attainment

	At least Primary (share of adult pop.)		At least Middle (share of adult pop.)		At least Secondary (share of adult pop.)	
	Land owners	*Landless*	*Land owners*	*Landless*	*Land owners*	*Landless*
Below canal	0.019***	0.009***	0.020***	0.009***	0.017***	0.005**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	0.002
Control group mean	0.521	0.435	0.356	0.271	0.235	0.002
Observations	90,042	89,715	90,042	89,715	90,042	89,715
R^2	0.600	0.480	0.580	0.470	0.550	0.420

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

Notes: Regression discontinuity results following Equation 5.1 for land owners and the landless. Panel A summarizes differences in the share of land owners, the size of land owners' plots, and the overall consumption of land owners and the landless. Panel B shows the consumption of landowners by quintile of land holding size. The bottom (1^{st}) quintile are the landowners with plots in the 0-20% range of the national distribution while the top (5^{th}) quintile are those landowners with total land holdings in the top 80-100% of the national distribution. The quintile break points in ascending order are 1.32, 2.30, 3.65, and 6.10 acres.

Table 5: Comparison to distant municipalities

Panel A. Irrigation outcomes

	Irrigated area (share)		Canal irrigated area (share)		Tubewell irrigated area (share)		Other irrigated area (share)	
	Treat	*Control*	*Treat*	*Control*	*Treat*	*Control*	*Treat*	*Control*
Group	0.064***	-0.002	0.083***	-0.003	0.003	0.004	-0.012	-0.003
	(0.027)	(0.020)	(0.044)	(0.026)	(0.022)	(0.021)	(0.020)	(0.015)
Control group mean	0.418		0.045		0.209		0.170	
Observations	252,262		251,772		252,040		250,394	
R^2	0.660		0.350		0.520		0.560	

Panel B. Agriculture outcomes

	Agricultural land (share)		Kharif ag. prod EVI-based (log)		Rabi ag. prod EVI-based (log)			
	Treat	*Control*	*Treat*	*Control*	*Treat*	*Control*		
Group	0.030***	0.004	-0.013	0.002	0.017	-0.031*		
	(0.022)	(0.022)	(0.025)	(0.020)	(0.035)	(0.028)		
Control group mean	0.595		7.738		7.244			
Observations	266,134		279,352		278,800			
R^2	0.560		0.710		0.620			

Panel C. Economic outcomes

	Population density (log)		Total emp. (share)		Services emp. (share)		Manuf. emp (share)	
	Treat	*Control*	*Treat*	*Control*	*Treat*	*Control*	*Treat*	*Control*
Group	0.111***	0.025	-0.006	-0.003	-0.005	-0.002	0.001	0.002
	(0.020)	(0.019)	(0.043)	(0.028)	(0.053)	(0.032)	(0.030)	(0.024)
Control group mean	5.616		0.077		0.053		0.018	
Observations	247,406		246,945		248,411		248,743	
R^2	0.420		0.010		0.000		0.020	

Panel D. Living Standards outcomes

	Consumption (log) total		Consumption (log) land holders		Consumption (log) landless		Literacy (share)	
	Treat	*Control*	*Treat*	*Control*	*Treat*	*Control*	*Treat*	*Control*
Group	0.020***	0.006	0.040***	0.014**	0.000	-0.002	0.020***	0.010***
	(0.020)	(0.020)	(0.019)	(0.019)	(0.018)	(0.018)	(0.020)	(0.021)
Control group mean	9.679		9.752		9.557		0.561	
Observations	258,638		248,156		245,790		271,985	
R^2	0.480		0.490		0.390		0.540	

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

Notes: Regression results following Equation 5.3 comparing treatment and control municipalities to distant municipalities far from the canal. Distant municipalities are defined as municipalities more than 15km away from a canal. The coefficients on the dummy variables for being in the treatment (below canal) or control (above canal) groups are reported here.

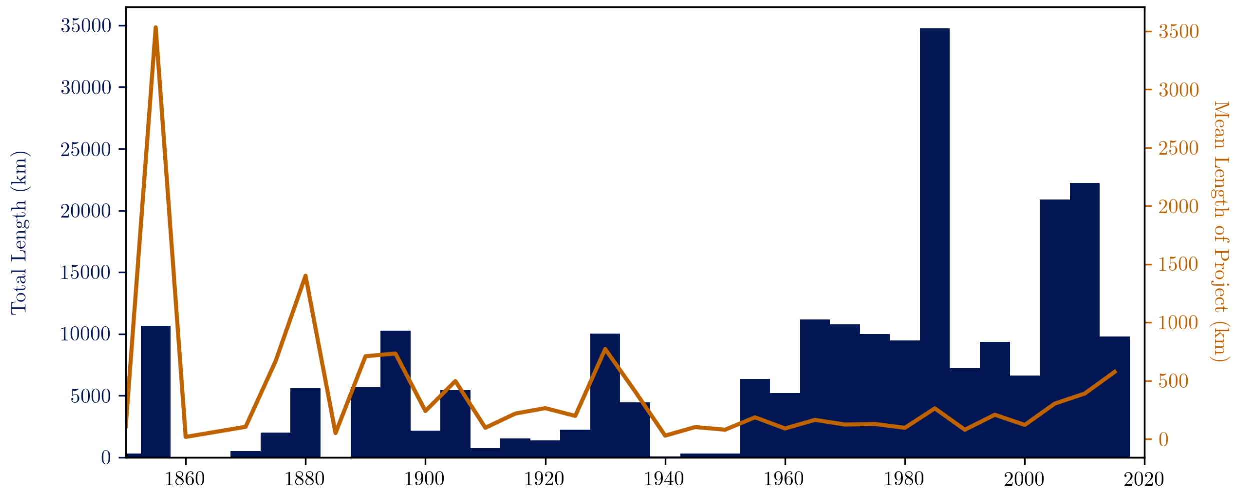
Table 6: Effect of canals on urbanization (event study)

	Town Existence (pop. 5,000)		Town Existence (pop. 10,000)		Population (log)	
	1	2	3	4	5	6
Any command area in town catchment area	0.014*		0.022**		0.026**	
	(0.007)		(0.009)		(0.011)	
Share of town catchment area in command area		0.066***		0.159***		0.193***
		(0.023)		(0.032)		(0.037)
Lagged Population (log)					0.587***	0.585***
					(0.008)	(0.008)
Constant	0.338***	0.332***	0.224***	0.204***	3.563***	3.550***
	(0.003)	(0.004)	(0.003)	(0.006)	(0.062)	(0.063)
N	88,632	88,632	88,632	88,632	81,246	81,246
R^2	0.70	0.70	0.64	0.64	0.88	0.88

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

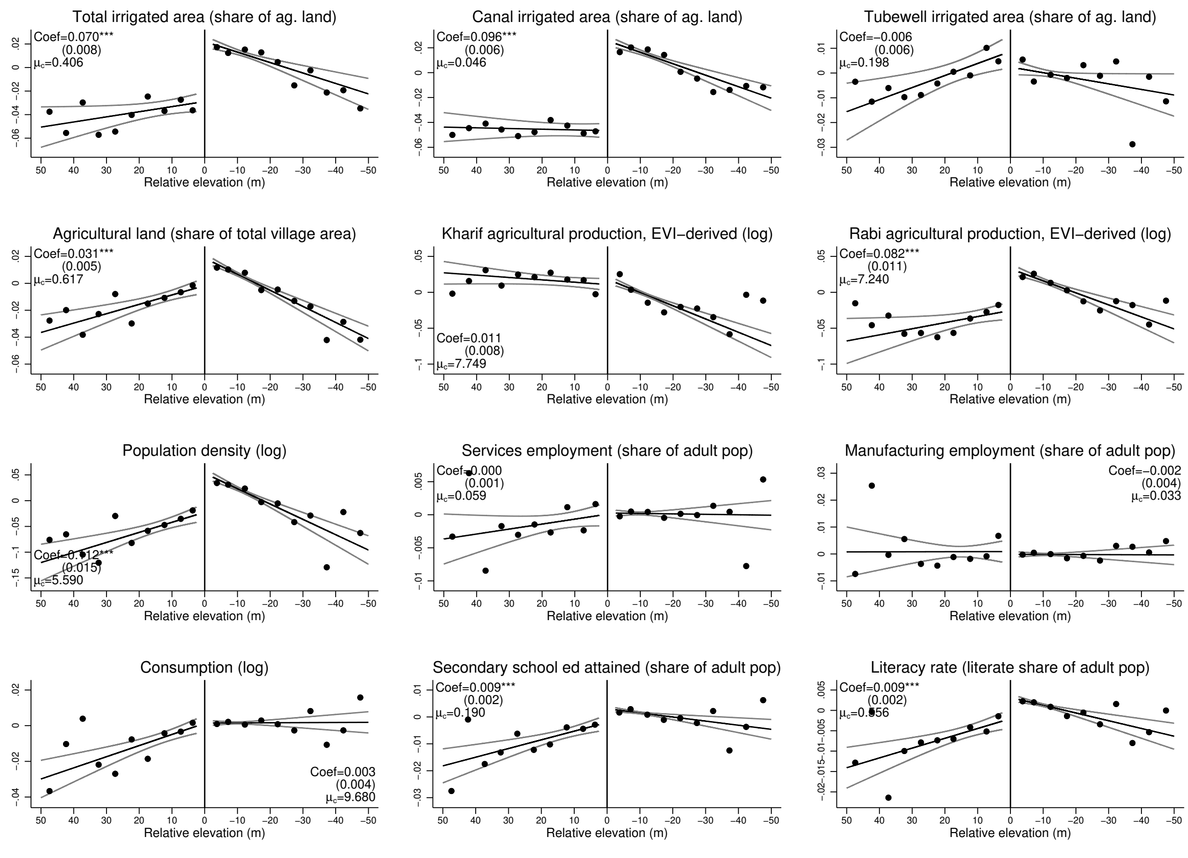
Notes: This table shows the estimated effect of canal area on town growth as estimated in Equation 5.4. Each column reports the β_1 values for various outcome variables. The outcome variable in columns 1 and 2 is the existence of a town with population 5,000 or greater as the outcome variable. In columns 3 and 4 the outcome variable is towns with population 10,000 or greater. Columns 5 and 6 use log population as the outcome variable, assuming that before towns appear in the time series they have a population of 2,000.

Figure 1: Canal construction through time



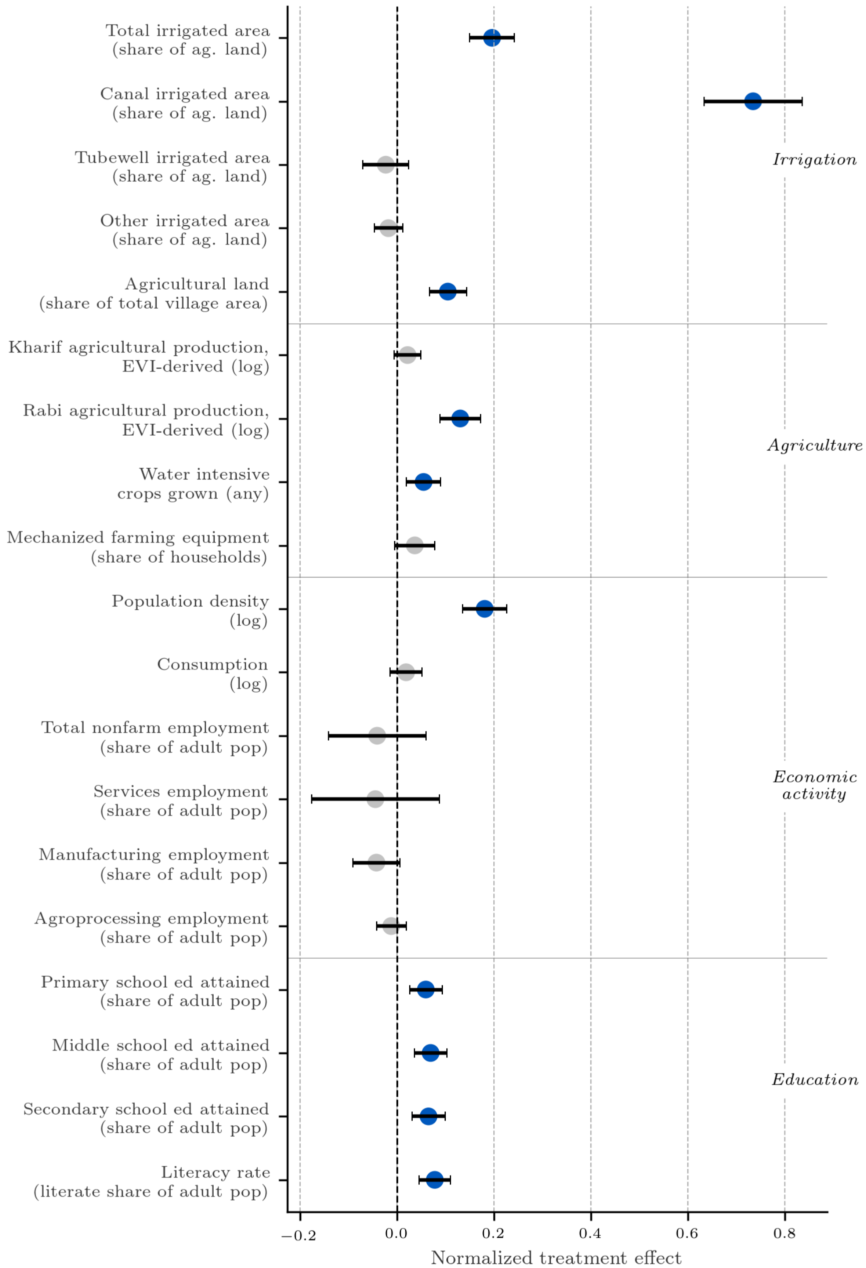
Notes: The total length of medium and major canals constructed in India from 1850-2013. The average length of a project in each 5-year time window is also plotted. Any canals with dates older than 1850 are coded as 1850 while any canals not completed before 2013 are not included. Note that 217 of the 1442 total canal projects reported, or 9% of total canal length in the canal shapefile, have an unknown date of completion and are not included in this plot. Additionally, 236 projects totaling 22% of total canal length in the data were not completed as of 2013 (the last date of our major outcome variables) and so are not included in this plot.

Figure 2: Regression discontinuity binscatters for key variables



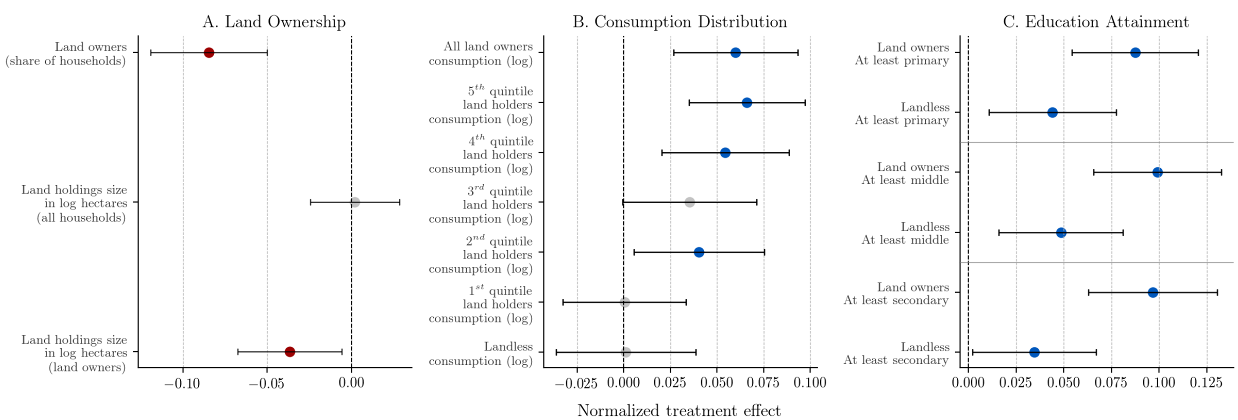
Notes: Each figure shows the binned scatterplot relationship between an outcome of interest and the RDD running variable (elevation relative to the nearest canal), after residualizing on the controls (ruggedness and rainfall) and subdistrict fixed effects. Treatment municipalities that lie below the canal have negative relative elevation while untreated municipalities that lie above the canal have positive relative elevation. All regressions were done following Equation 5.1. The regression discontinuity coefficient (Coef) for each variable is reported with stars indicating the significance and the standard error in parentheses below. The control group mean is also reported (μ_c).

Figure 3: Regression discontinuity coefficients for main outcomes



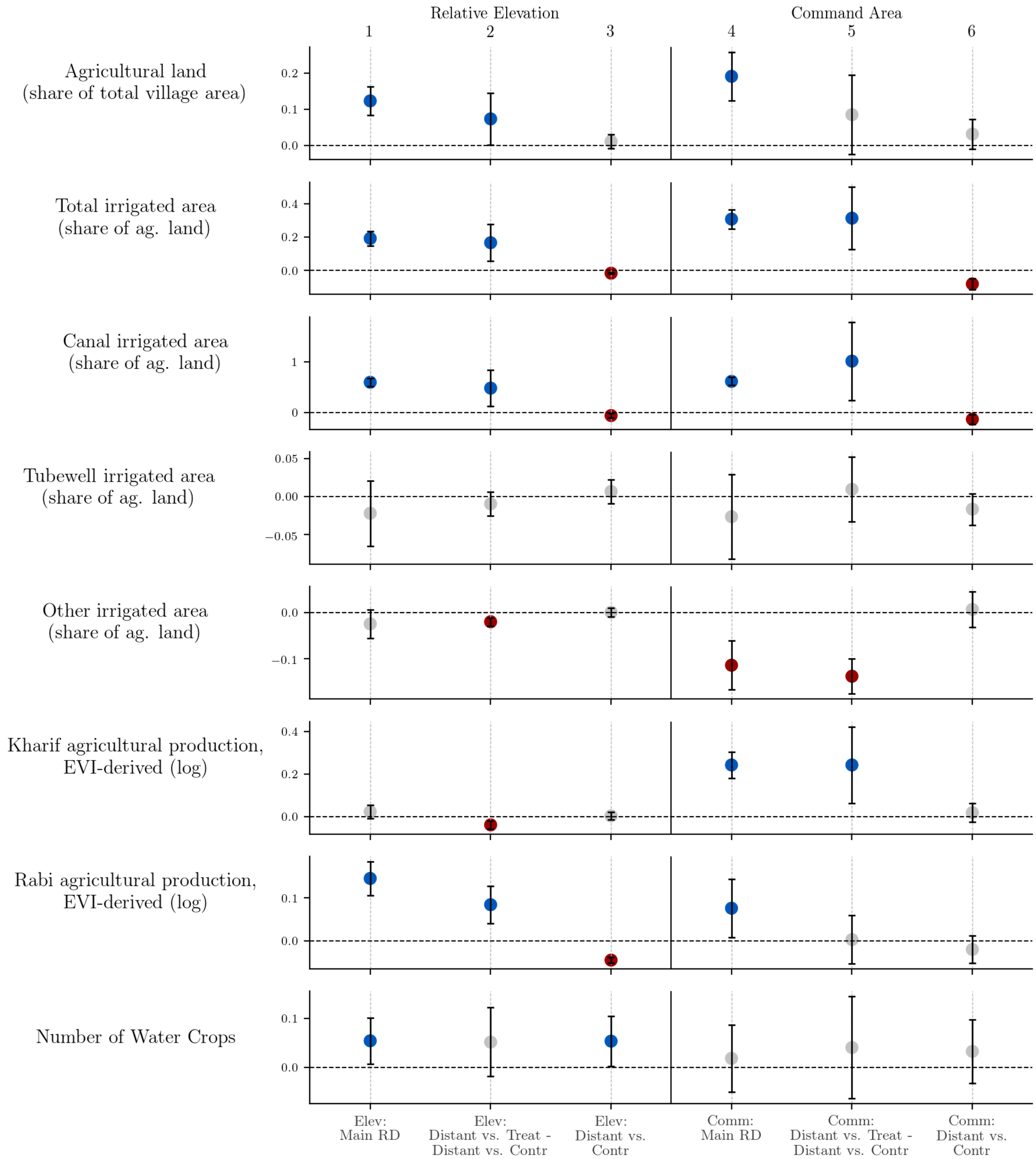
Notes: This figure shows the regression discontinuity results for our main outcomes using our preferred specification and following Equation 5.1. Blue points indicate normalized treatment effects above 0, red below zero, and gray indicates an insignificant result. The normalized treatment effect is calculated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

Figure 4: Land ownership outcomes



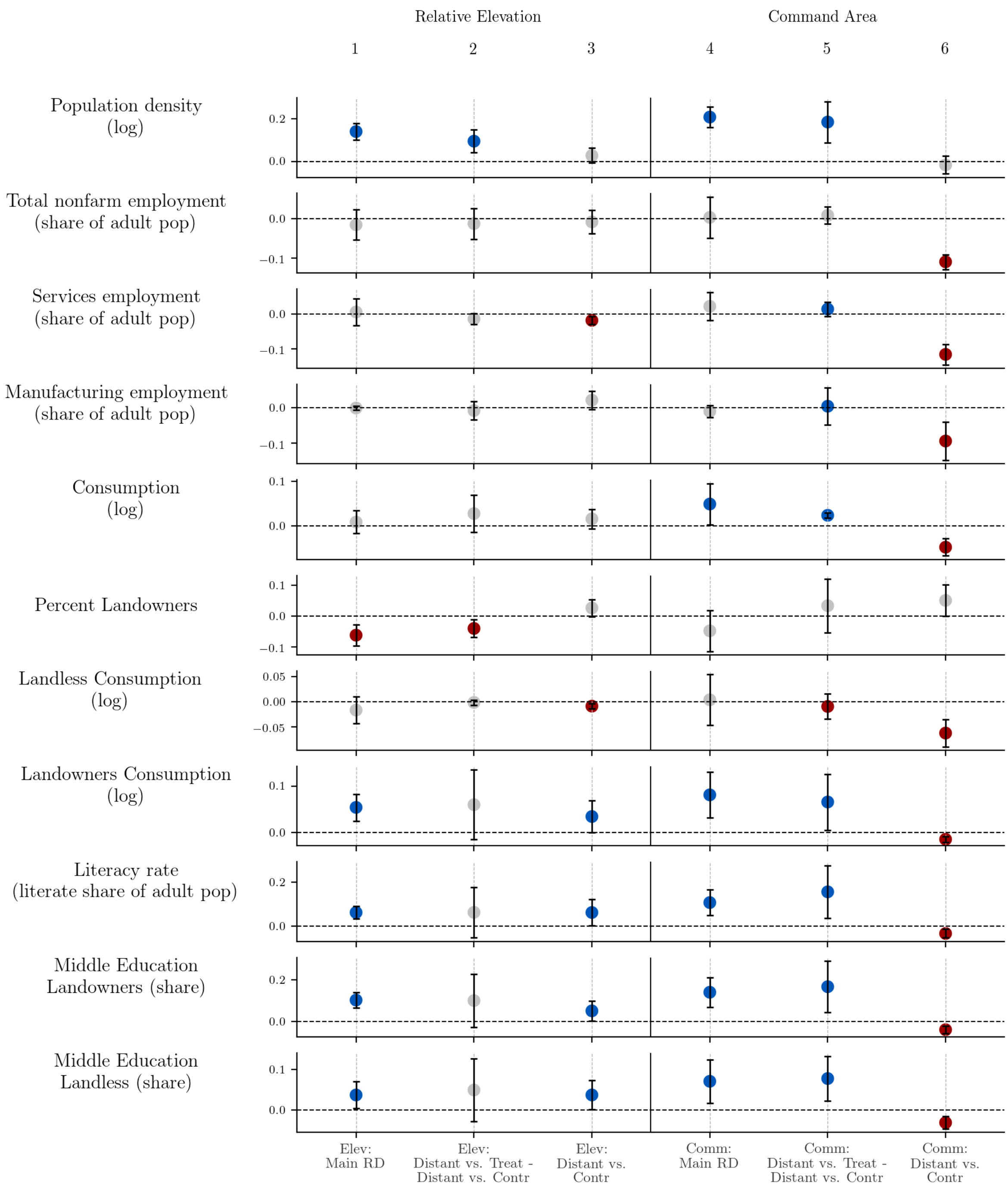
Notes: This figure shows the regression discontinuity coefficients for outcomes disaggregated by land ownership. Following Equation 5.1, we estimate the RD coefficients for outcomes based on land ownership status. Error bars indicate the 95% confidence interval for each estimate.

Figure 5: RDD and alternate control group analysis results for irrigation and agricultural outcomes



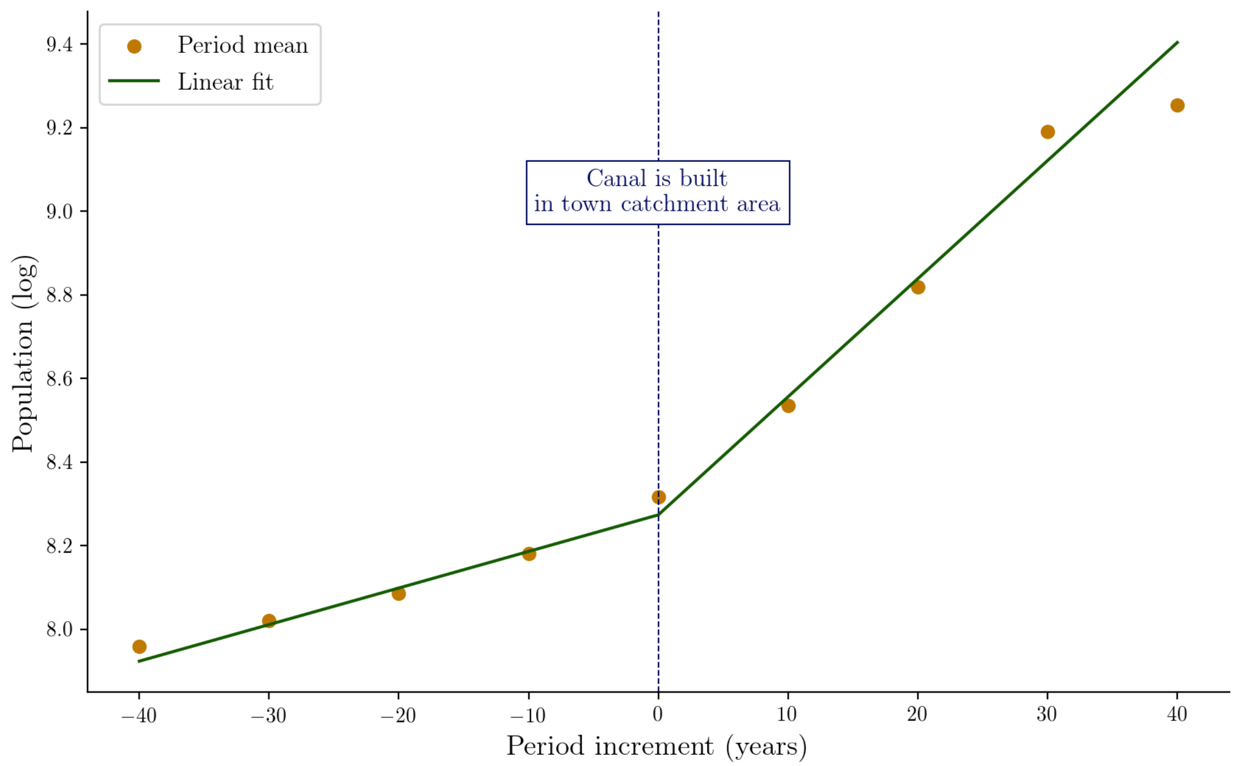
Notes: This figure combines all village-level results for agricultural and irrigation outcomes. Blue points indicate positive, significant coefficients while red points indicate negative, significant coefficients. Grey points are insignificant. The left panel shows results using the relative elevation strategy while the right panel shows results using the command area strategy. Columns 1 and 4 report the main RDD result of each outcome variable as reported in Table 3. Columns 2 and 5 report the difference between the coefficient on Treatment and Control municipalities as calculated by Equation 5.3. This value is equivalent to the difference between the Treat and Control columns in Table 5. Columns 3 and 6 report the Control column from Table 5, the effect on control municipalities relative to distant municipalities.

Figure 6: RDD and alternate control group analysis results for economic and education outcomes



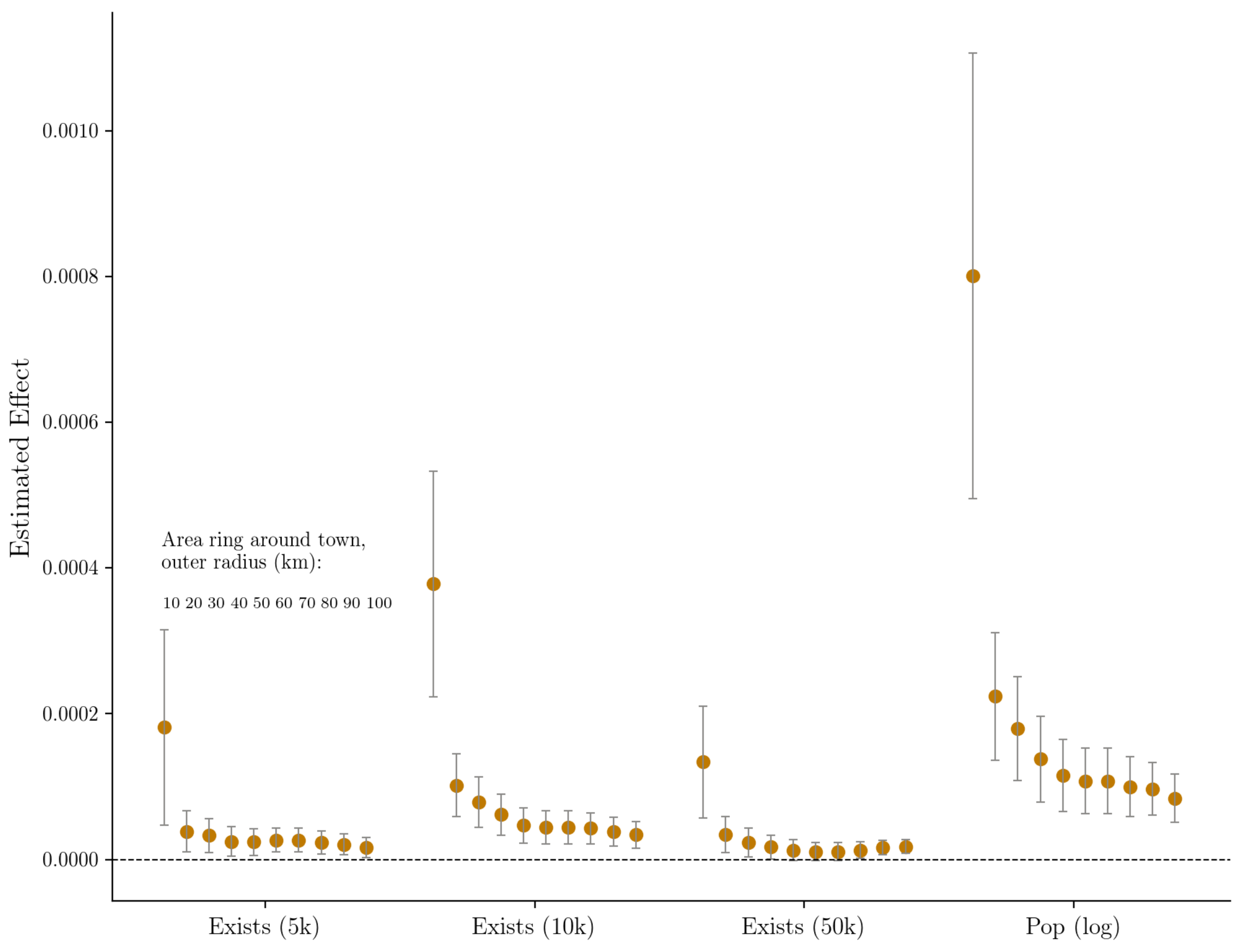
Notes: This figure combines all village-level results for economic and education outcomes. Blue points indicate positive, significant coefficients while red points indicate negative, significant coefficients. Grey points are insignificant. The left panel shows results using the relative elevation strategy while the right panel shows results using the command area strategy. Columns 1 and 4 report the main RDD result of each outcome variable as reported in Table 3. Columns 2 and 5 report the difference between the coefficient on Treatment and Control municipalities as calculated by Equation 5.3. This value is equivalent to the difference between the Treat and Control columns in Table 5. Columns 3 and 6 report the Control column from Table 5, the effect on control municipalities relative to distant municipalities.

Figure 7: Trend break in town population growth after canal construction



Notes: This figure shows the trend break that occurs in town population after canal construction. Towns are aligned by period, where period 0 is the decade in which a command area first appeared in the town catchment area, with the 30km town radius used in this figure. Period increment 10 indicates 1 decade after the command area appearance while period increment -10 indicates 1 decade before the first command area appearance.

Figure 8: Difference in Difference, town population



Notes: This figure shows the effect calculated as β_1 in Equation 5.4. Each point shows an individual estimate with its 95% confidence interval. The binary outcome of a town existing, using 5,000, 10,000, and 50,000 population thresholds, as well as the log of the town population are all estimated as the outcome variable $y_{i,t}$. Results are clustered by these outcome variables along the x-axis, with each individual estimate representing a different area ring around the town with and outer radius r ranging from 10 to 100km, and inner radius $10-r$.

References

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.

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

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 3 2020, *110*, 797–823.

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

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

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

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural Productivity and Structural Transformation: Evidence from Brazil," *American Economic Review*, 6 2016, *106*, 1320–1365.

___ **, Gabriel Garber, and Jacopo Ponticelli**, "Capital Accumulation and Structural Transformation," *The Quarterly Journal of Economics*, 5 2020, *135*, 1037–1094.

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

Douie, James, "The Punjab Canal Colonies," *Journal of the Royal Society of Arts*, 5 1914, *62*, 611–623.

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

Emerick, Kyle, "Agricultural productivity and the Sectoral Reallocation of Labor in Rural India," *Journal of Development Economics*, 11 2018, *135*, 488–503.

Foster, Andrew D. and Mark R. Rosenzweig, "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review*, 9 1996, *86*, 931–953.

___ **and** ___ , "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970â2000*," *Economic Development and Cultural Change*, 7 2004, *52*, 509–542.

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

___ **and** ___ , "Chapter 47 Economic Development and the Decline of Agricultural Employment," *Handbook of Development Economics*, 1 2007, *4*, 3051–3083.

___ **and** ___ , "Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size," *NBER Working Paper*, 10 2017.

Gollin, Douglas, Casper Worm Hansen, and Asger Mose Wingender, "Two Blades of Grass: The Impact of the Green Revolution," *Journal of Political Economy*, 6 2021, *129.*

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

___ **, Stephen Parente, and Richard Rogerson**, "The Role of Agriculture in Development," *American Economic Review*, 5 2002, *92*, 160–164.

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

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

Jones, Maria, Florence Kondylis, John Loeser, and Jeremy Magruder, "Factor Market Failures and the Adoption of Irrigation in Rwanda," *National Bureau of Economic Research*, 1 2020.

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

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

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

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

Moscona, Jacob, "Agricultural Development and Structural Change, Within and Across Countries *," 2020.

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

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

Ngai, Rachel and Christopher A. Pissarides, "Structural Change in a Multi-sector Model of Growth," 2004.

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

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

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

——, *Transforming Traditional Agriculture* Studies in comparative economics ; 3, New Haven ; London: Yale University Press, 1964.

Shah, Manisha and Bryce Millett Steinberg, "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital," *Journal of Political Economy (forthcoming)*, 2016, *No. 19140.*

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

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

Table A1: Additional RDD outcomes

	Municipality is a Town (likelihood)
Panel A: Relative elevation	
Below canal	0.010
	(0.006)
Control group mean	0.023
Observations	97,879
R^2	0.150
Panel B: Command area	
Inside command area	0.008
	(0.011)
Control group mean	
Observations	49,010
R^2	0.210

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

Notes: Additional outcomes reported for both the relative elevation and commanda rea boundary specifications. These results use the balanced sample for each specification.

Table A2: Regression discontinuity results for irrigation outcomes for all specifications

	Total irrigated area (share of ag. land)	Canal irrigated area (share of ag. land)	Tubewell irrigated area (share of ag. land)	Other irrigated area (share of ag. land)
Panel A: Relative elevation, full sample				
Below canal	0.061***	0.084***	-0.006	-0.008***
	(0.001)	(0.001)	(0.000)	(0.000)
Control group mean	0.434	0.068	0.212	0.159
Observations	243,510	243,186	243,319	241,538
R^2	0.640	0.450	0.480	0.570
Panel B: Command area, full sample				
Below canal	0.071***	0.108***	-0.010**	-0.015***
	(-0.005)	(-0.006)	(0.000)	(0.002)
Control group mean	0.412	0.054	0.212	0.153
Observations	232,601	232,062	232,354	230,738
R^2	0.670	0.450	0.540	0.690
Panel C: Relative elevation, donut hole				
Below canal	0.075***	0.104***	-0.007	-0.009**
	(0.001)	(0.002)	(0.000)	(0.000)
Control group mean	0.387	0.050	0.180	0.163
Observations	133,195	132,941	133,046	131,567
R^2	0.600	0.390	0.490	0.640
Panel D: Command area, donut hole				
Below canal	0.133***	0.193***	-0.025**	-0.022**
	(-0.001)	(-0.001)	(-0.002)	(0.001)
Control group mean	0.484	0.060	0.299	0.138
Observations	57,540	57,393	57,462	56,934
R^2	0.710	0.500	0.570	0.530
Panel E: Relative elevation, balanced sample				
Below canal	0.069***	0.092***	-0.006	-0.006
	(0.001)	(0.001)	(0.000)	(0.000)
Control group mean	0.418	0.045	0.209	0.170
Observations	89,628	89,467	89,539	88,757
R^2	0.610	0.380	0.480	0.640
Panel F: Command area, balanced sample				
Below canal	0.120***	0.155***	-0.004	-0.021*
	(-0.001)	(-0.002)	(-0.002)	(0.002)
Control group mean	0.534	0.063	0.332	0.156
Observations	43,342	43,267	43,311	42,850
R^2	0.730	0.510	0.570	0.520

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

Notes: The full sample, reported in panels A and B, contains all municipalities ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation for the relative elevation strategy. For the command area strategy, municipalities must be within 25km of the command area boundary. The donut hole sample drops all municipalities 0-2.5m from the canal for the relative elevation strategy (Panel C) or within 2.5km of the command area boundary for the command area strategy (Panel D). The balanced sample, shown in panels E and F, then imposes a balance criteria on ruggedness by dropping all subdistricts in which the there is a ≥25% difference in average ruggedness between treatment and control municipalities for the elevation specification, and all 10km segment groups with ≥25% difference for the command area specification. Panel E is our preferred specification.

Table A3: Regression discontinuity results for agricultural outcomes for all specifications

	Agricultural land (share of village area)	Kharif (monsoon) ag. prod (log)	Rabi (winter) ag. prod (log)	Water intensive crops (any)	Mechanized farm equip. (share of households)
Panel A: Relative elevation, full sample					
Below canal	0.032***	0.025***	0.051***	0.028***	0.003**
	(0.002)	(0.003)	(0.002)	(0.001)	(0.000)
Control group mean	0.593	7.744	7.268	0.697	0.043
Observations	255,083	257,253	255,368	203,904	244,668
R^2	0.580	0.790	0.710	0.720	0.340
Panel B: Command area, full sample					
Below canal	0.047***	0.055***	0.057***	0.017***	0.003***
	(-0.003)	(-0.015)	(-0.005)	(-0.002)	(0.000)
Control group mean	0.581	7.700	7.312	0.675	0.040
Observations	245,872	248,131	247,414	198,080	235,445
R^2	0.590	0.800	0.700	0.780	0.300
Panel C: Relative elevation, donut hole					
Below canal	0.041***	0.023***	0.068***	0.028***	0.006***
	(0.002)	(0.003)	(0.002)	(0.000)	(0.000)
Control group mean	0.565	7.735	7.234	0.672	0.039
Observations	141,755	142,862	142,378	109,864	135,980
R^2	0.580	0.810	0.680	0.740	0.310
Panel D: Command area, donut hole					
Below canal	0.048***	0.125***	0.100***	0.031	0.006**
	(-0.001)	(-0.003)	(0.001)	(-0.002)	(0.000)
Control group mean	0.624	7.639	7.378	0.755	0.043
Observations	63,536	64,153	64,010	54,067	60,573
R^2	0.690	0.790	0.770	0.800	0.350
Panel E: Relative elevation, balanced sample					
Below canal	0.028***	0.014	0.074***	0.027***	0.004*
	(0.002)	(0.003)	(0.002)	(0.001)	(0.000)
Control group mean	0.595	7.738	7.244	0.648	0.047
Observations	96,403	97,089	96,792	75,349	92,357
R^2	0.610	0.820	0.700	0.720	0.310
Panel F: Command area, balanced sample					
Below canal	0.027**	0.147***	0.061**	0.028	0.008**
	(0.000)	(-0.003)	(0.001)	(-0.002)	(0.000)
Control group mean	0.655	7.610	7.389	0.783	0.047
Observations	48,307	48,811	48,699	41,788	46,124
R^2	0.720	0.800	0.790	0.770	0.350

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

Notes: The full sample, reported in panels A and B, contains all municipalities $\leq$10km from the nearest canal in distance and $\pm$50m from the nearest canal in elevation for the relative elevation strategy. For the command area strategy, municipalities must be within 25km of the command area boundary. The donut hole sample drops all municipalities 0-2.5m from the canal for the relative elevation strategy (Panel C) or within 2.5km of the command area boundary for the command area strategy (Panel D). The balanced sample, shown in panels E and F, then imposes a balance criteria on ruggedness by dropping all subdistricts in which the there is a $\geq$25% difference in average ruggedness between treatment and control municipalities for the elevation specification, and all 10km segment groups with $\geq$25% difference for the command area specification. Panel E is our preferred specification.

Table A4: Regression discontinuity results for economic outcomes for all specifications

	Population density (log)	Total emp. (share of adult pop.)	Services emp. (share of adult pop.)	Manuf. emp (share of adult pop.)	Consumption per capita (log)
Panel A: Relative elevation, full sample					
Below canal	0.100***	-0.004	-0.002	-0.002**	0.009***
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	5.690	0.083	0.056	0.019	9.662
Observations	235,840	237,477	238,355	238,745	244,570
R^2	0.510	0.020	0.010	0.060	0.530
Panel B: Command area, full sample					
Below canal	0.139***	0.001	0.003	-0.001	0.019***
	(-0.011)	(0.001)	(0.001)	(0.000)	(-0.002)
Control group mean	5.712	0.081	0.055	0.018	9.646
Observations	226,560	227,426	228,553	228,870	235,379
R^2	0.520	0.060	0.020	0.130	0.550
Panel C: Relative elevation, donut hole					
Below canal	0.144***	-0.002	-0.001	-0.001	0.014***
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
Control group mean	5.567	0.083	0.056	0.019	9.654
Observations	133,180	132,703	133,264	133,449	135,943
R^2	0.440	0.010	0.010	0.050	0.530
Panel D: Command area, donut hole					
Below canal	0.190***	0.016	0.013	0.001	0.019
	(0.001)	(0.003)	(0.003)	(0.000)	(-0.001)
Control group mean	6.021	0.078	0.053	0.018	9.645
Observations	56,403	58,777	59,033	59,116	60,551
R^2	0.640	0.020	0.010	0.080	0.570
Panel E: Relative elevation, balanced sample					
Below canal	0.131***	-0.005	-0.004	-0.003*	0.005
	(0.004)	(-0.001)	(0.000)	(0.000)	(0.000)
Control group mean	5.616	0.077	0.053	0.018	9.679
Observations	90,223	90,695	91,022	91,126	92,327
R^2	0.460	0.010	0.010	0.040	0.530
Panel F: Command area, balanced sample					
Below canal	0.175***	0.017	0.016	0.002	0.017*
	(0.002)	(0.003)	(0.003)	(0.000)	(-0.001)
Control group mean	6.186	0.070	0.051	0.019	9.664
Observations	42,402	44,895	45,005	45,071	46,104
R^2	0.640	0.020	0.020	0.050	0.560

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

Notes: The full sample, reported in panels A and B, contains all municipalities ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation for the relative elevation strategy. For the command area strategy, municipalities must be within 25km of the command area boundary. The donut hole sample drops all municipalities 0-2.5m from the canal for the relative elevation strategy (Panel C) or within 2.5km of the command area boundary for the command area strategy (Panel D). The balanced sample, shown in panels E and F, then imposes a balance criteria on ruggedness by dropping all subdistricts in which the there is a ≥30% difference in average ruggedness between treatment and control municipalities for the elevation specification, and all 10km segment groups with ≥25% difference for the command area specification. Panel E is our preferred specification.

Table A5: Regression discontinuity results for education outcomes for all specifications

	At least primary (share of adult pop.)	At least middle (share of adult pop.)	At least secondary (share of adult pop.)	Literacy (share of pop.)
Panel A: Relative elevation, full sample				
Below canal	0.015***	0.015***	0.011***	0.008***
	(0.001)	(0.000)	(0.000)	(0.000)
Control group mean	0.464	0.309	0.187	0.554
Observations	244,447	244,447	244,447	258,854
R^2	0.580	0.560	0.520	0.590
Panel B: Command area, full sample				
Below canal	0.015***	0.015***	0.013***	0.009***
	(-0.002)	(-0.002)	(-0.002)	(-0.002)
Control group mean	0.454	0.305	0.183	0.549
Observations	235,267	235,267	235,267	249,890
R^2	0.610	0.580	0.540	0.620
Panel C: Relative elevation, donut hole				
Below canal	0.021***	0.020***	0.015***	0.013***
	(0.001)	(0.000)	(0.000)	(0.000)
Control group mean	0.450	0.297	0.179	0.547
Observations	135,862	135,862	135,862	143,990
R^2	0.590	0.580	0.540	0.600
Panel D: Command area, donut hole				
Below canal	0.025***	0.024***	0.023***	0.019***
	(-0.001)	(-0.001)	(0.000)	(0.000)
Control group mean	0.450	0.304	0.176	0.548
Observations	60,518	60,518	60,518	64,416
R^2	0.640	0.610	0.570	0.670
Panel E: Relative elevation, balanced sample				
Below canal	0.012***	0.012***	0.009***	0.010***
	(0.001)	(0.000)	(0.000)	(0.000)
Control group mean	0.469	0.312	0.190	0.561
Observations	92,268	92,268	92,268	97,879
R^2	0.570	0.560	0.530	0.590
Panel F: Command area, balanced sample				
Below canal	0.021***	0.021***	0.021***	0.019***
	(-0.001)	(-0.001)	(0.000)	(0.000)
Control group mean	0.471	0.322	0.187	0.562
Observations	46,081	46,081	46,081	48,969
R^2	0.650	0.620	0.570	0.680

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

Notes: The full sample, reported in panels A and B, contains all municipalities ≤10km from the nearest canal in distance and ±50m from the nearest canal in elevation for the relative elevation strategy. For the command area strategy, municipalities must be within 25km of the command area boundary. The donut hole sample drops all municipalities 0-2.5m from the canal for the relative elevation strategy (Panel C) or within 2.5km of the command area boundary for the command area strategy (Panel D). The balanced sample, shown in panels E and F, then imposes a balance criteria on ruggedness by dropping all subdistricts in which the there is a ≥25% difference in average ruggedness between treatment and control municipalities for the elevation specification, and all 10km segment groups with ≥25% difference for the command area specification. Panel E is our preferred specification.

Table A6: Sensitivity to bandwidth and canal distance

Panel A. Regression discontinuity bandwidth

Bandwidth (m)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
15	0.064***	0.037**	0.077***	0.005*	0.124	81,389
	(0.010)	(0.016)	(0.024)	(0.003)	(0.146)	
25	0.064***	0.057***	0.100***	0.009***	-0.050	95,863
	(0.008)	(0.012)	(0.018)	(0.002)	(0.059)	
50	0.068***	0.080***	0.131***	0.010***	-0.044	98,553
	(0.008)	(0.012)	(0.017)	(0.002)	(0.055)	
75	0.072***	0.074***	0.139***	0.010***	-0.084	98,572
	(0.008)	(0.012)	(0.017)	(0.002)	(0.054)	

Panel B. Percent difference in ruggedness

Percent difference in ruggedness (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
10%	0.063***	0.056***	0.119***	0.007**	-0.030	58,480
	(0.011)	(0.016)	(0.022)	(0.003)	(0.039)	
25%	0.069***	0.074***	0.131***	0.010***	-0.044	98,553
	(0.008)	(0.012)	(0.017)	(0.002)	(0.055)	
50%	0.072***	0.070***	0.138***	0.011***	-0.209***	126,371
	(0.007)	(0.011)	(0.015)	(0.002)	(0.052)	

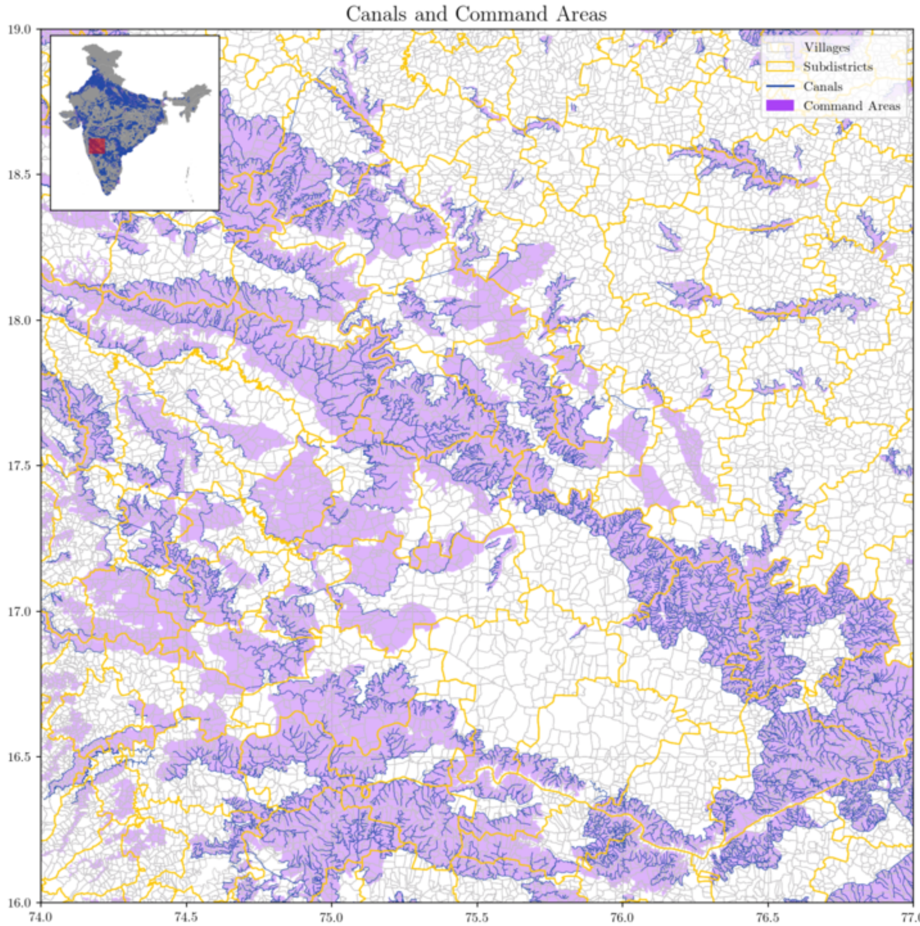
Panel C. Distance to Canal

Max distance to canal (km)	Total irrigated area (share of ag. land)	Rabi (winter) ag. prod (log)	Population density (log)	Literacy (share of pop.)	Balance (ruggedness)	Sample size
5	0.080***	0.071***	0.155***	0.008***	-0.034	64,307
	(0.011)	(0.016)	(0.021)	(0.003)	(0.053)	
10	0.069***	0.074***	0.131***	0.010***	-0.044	98,553
	(0.008)	(0.012)	(0.017)	(0.002)	(0.055)	
15	0.066***	0.077***	0.117***	0.002	-0.066	121,636
	(0.007)	(0.011)	(0.015)	(0.007)	(0.042)	

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

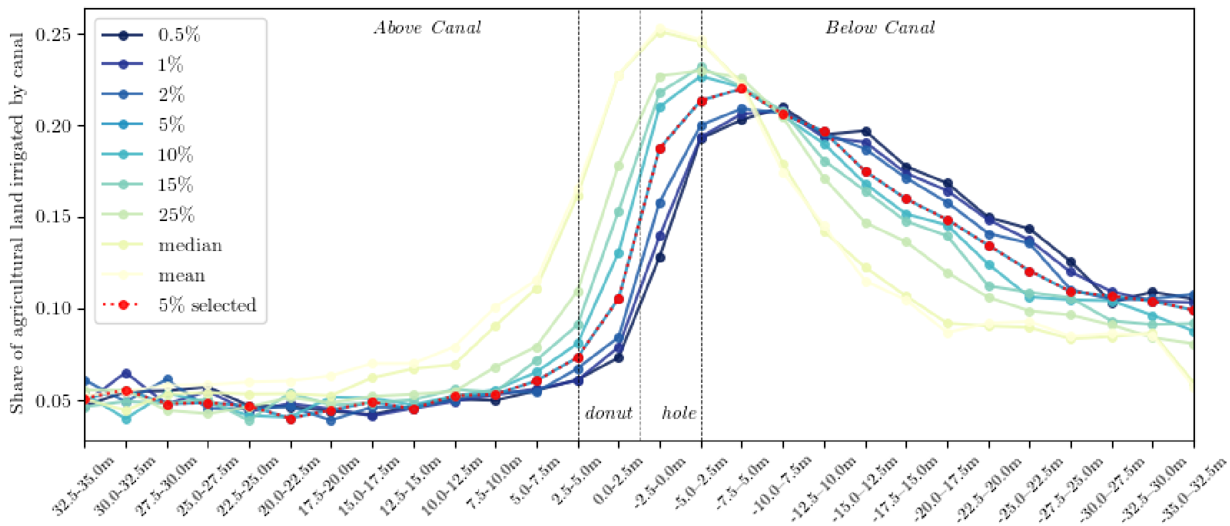
Notes: This table shows the sensitivity of our major results to changes in the construction of our sample. Each panel shows the results if one assumption were to be changed. The bolded parameters indicate the values used in our preferred sample. These preferred values are used for the two parameters not being tested in each panel. In panel A, we modify the bandwidth of the regression discontinuity where 50m would include municipalities that lie 50m above to 50m below the nearest canal. Here we test 15m, 25m, and 75m bandwidths in addition to our preferred 50m bandwidth. In panel B, we modify the threshold allowed for the average difference in ruggedness between treatment and control municipalities for each fixed effect group. We test 10% (more strict) and 50% (less strict) in addition to our preferred 25% threshold. Lastly in panel C we modify the maximum distance a village may lie away from the nearest canal. Here we test 5km and 15km in addition to our preferred 10km.

Figure A1: Illustrating village assignment by command areas.



Notes: This map shows a region in Southwestern India to illustrate the command area data. Village polygons are in gray, with subdistricts outlined in yellow. The command areas are in purple with the canals themselves are the blue lines. Municipalities that fall within the command area are considered treated, while those outside are the control municipalities.

Figure A2: Calculating the relative elevation of each village.



Notes: Each line in this figure uses a different moment of the distribution of elevation in a village polygon to define the relative elevation between that village 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 municipalities below the canal. We select the 5^{th} percentile to define village elevation.