The Cost of Distance: Geography and Governance in Rural India*
PRELIMINARY: PLEASE DO NOT CITE WITHOUT PERMISSION

Sam Asher† Karan Nagpal‡ Paul Novosad§

February 21, 2017

Abstract

Spatial inequalities are severe in developing countries, particularly in terms of access to public goods and services. We show that the geography of public administration contributes to this inequality. We construct a high-resolution spatial dataset on 600,000 Indian villages, with information on household income, assets, employment structure, public goods, and geographic location of administrative capitals. We exploit administrative boundaries that generate sharp jumps in distance to administrative capitals but not in market access, population density or distance to highways. Villages that are more distant from administrative capitals receive fewer paved roads, have lower literacy and more limited participation in non-agricultural activities. We also document higher costs of infrastructure provision in remote areas, which may help to explain lower provision.

JEL Codes: R12, D63, H41, O18

*This version January 2017. We thank seminar participants at the Center for Global Development Washington DC, the DFID-IZA Workshop in Oxford, the Indian Statistical Institute annual conference in Delhi, and at the Centre for the Study of African Economies Oxford. We are indebted to Taewan Roh and Kathryn Nicholson for exemplary research assistance.

†World Bank Research Group

‡University of Oxford, karan.nagpal@economics.ox.ac.uk

§Dartmouth College

I Introduction

Despite large advances in communications and information technology, physical distance from large markets remains an important determinant of economic outcomes and a major detriment to economic development (Fafchamps and Shilpi, 2003; Feyrer, 2009; Storeygard, 2016; Atkin and Donaldson, 2015). Distance from markets contributes to spatial inequality, which is particularly severe in developing countries (Kanbur and Venables, 2005; Kanbur and Rapoport, 2005; Moretti, 2011; Bryan and Morten, 2015). For example, average household consumption in richer regions of developing countries can be almost 75% higher than in poorer regions of the same country; the corresponding differential for developed countries is less than 25% (The World Bank, 2009). But this spatial inequality can also be very local. Our calculations using Indian data suggest that about 77% of the national variation in mean monthly earnings occurs within districts.[1] There is also substantial spatial inequality in access to public goods. For example, 31% percent of the world's rural population lives in settlements more than 2 kilometeres from a paved road (The World Bank, 2015). How governments choose to deliver public goods can have important implications for this unequal access (Bardhan, 2002), and perpetuate spatial poverty traps (Jalan et al., 1997).

In this paper, we provide causal evidence that a village's distance from its administrative capital, which we refer to as "administrative remoteness", has significant negative consequences for public goods provision and economic outcomes. To do this, we assemble a high-resolution spatial panel dataset covering approximately 600,000 Indian villages and 4,000 towns, with geographic location and information on public goods, average earnings, household assets and employment structure over a 21-year period (1991-2012).

To isolate the effect of administrative remoteness from other urban-rural channels, we implement a spatial regression discontinuity design that compares villages on either side of

[1] Districts are the third tier of administration in India, after the federal and provincial governments, equivalent to counties in the UK. Earnings data has been generated by combining the 68th round of the National Sample Survey (2011-12) with the Socioeconomic and Caste Census 2012 as described in Section II

administrative borders. These borders generate discontinuous jumps in distance to administrative capitals, but not in measures of access to markets such as distance to towns and cities, distance to highways, and local population density.

We find that an increase in "administrative remoteness", or distance to district capital in the Indian context[2], reduces the provision of those public goods that are provided by the district administration, such as paved roads, but does not affect public goods provided by a higher tier of administration, such as electricity. Increase in distance to district capital also affects economic outcomes adversely, reducing literacy rates and the proportion of rural workforce engaged in non-agricultural activities. For example, an increase of one standard deviation in distance to district capital (a change of about 24 kilometers) reduces the probability that the village has a paved road by 1.9 percentage points. It also reduces the proportion of workforce in nonfarm activities by 3.1 percentage points, and average literacy rates by 1 percentage point. Our results are robust to the bandwidth around the district borders, to the specific form of the geographic polynomial, and to dropping district borders that also happen to be state borders.

A potential mechanism for these results is that it is costlier to provide public goods and services to administratively remote areas, whether due to information or other frictions. We find evidence consistent with this hypothesis using construction cost and project duration data from a national rural roads program, the Pradhan Mantri Gram Sadak Yojna (PMGSY). An increase of one standard deviation in distance to district capital increases the cost per kilometer under the PMGSY program by 2.2 percentage points.

This paper contributes to several strands of the economics literature. There is a long-standing literature on the costs and benefits of decentralization (Bardhan, 2002). We see our

[2]Districts are the third tier of administration in India, after the federal and provincial governments. Several public programs in India, whether federal or state-level, are implemented by the district administration based in district capitals. There were 640 districts in India in 2011, with an average of approximately two million citizens per district.

concept of "administrative remoteness" as being the opposite of decentralization. Whereas decentralization is about bringing the "state" closer to its citizens (both in distance and hierarchy), "administrative remoteness" captures the effect of the state being more physically distant from its citizens. We provide empirical evidence that distance from district capitals matters for the provision of those public goods that are provided by the district administration, such as paved roads and secondary schools, and leads to inequality in their provision even within fairly small geographical areas. We also show that this distance does not matter for the provision of public goods that are provided by higher tiers of administration, such as electricity, whose provision is managed at the provincial level.

We also contribute to the literature documenting inequality in living standards across the world, especially as a function of geographic location. One of the dimensions that has received a lot of attention in economics is the urban-rural gap in consumption and living standards. For example, the urban-rural gap accounts for 40% of the average inequality in a sample of sixty developing countries (Young, 2013). In India, though the urban wage premium has declined from 59% in 1983, it was still a substantial 13% in 2010 (Hnatkovska and Lahiri, 2013). Our estimates suggest that the extent of inequality within rural areas, even in fairly narrow geographical areas, can be large.

Further, this paper adds to the literature on spatial gradients for governance and "state capacity". This literature has documented, for example, that African states get weaker as we move away from capital cities (Bates, 1983; Herbst, 2014; Michalopoulos and Papaioannou, 2014), and that even in more developed countries such as the United States, more isolated state capitals suffer from higher corruption and reduced accountability (Campante et al., 2014). Our work is closest in spirit to the descriptive work in Krishna and Schober (2014), which documents substantial spatial gradients in governance indicators in two districts in southern India. We find here that these spatial gradients represent a more general and pervasive phenomenon. We also provide causal evidence that the governance or "state ca-

pacity" gradients have a negative effect on a rich set of public goods and economic indicators, and hence contribute to the low living standards in rural parts of many developing countries.

The rest of the paper proceeds as follows: Section II describes our data and the construction of our main variables of interest. Section III explains the empirical strategy. Section IV presents and discusses our results, including robustness checks. Section V concludes.

II Data

In order to study the relationship between administrative remoteness and the rural economy, we construct a unique panel dataset on Indian villages covering a 21-year period (1991-2012). We use data from two waves of the Socioeconomic census (2002 and 2012) and three waves of the Population Census (1991, 2001 and 2011). We also obtain geocoordinates for all towns and villages in India and use these to calculate our distance measures. Below, we describe each source in greater detail.

II.A Socioeconomic census

The primary outcomes presented in this paper come from individual- and household-level microdata from a national socioeconomic census. Beginning in 1992, the Government of India has conducted multiple household censuses in order to determine eligibility for various government programs (Alkire and Seth, 2012). In 1992, 1997 and 2002, these were referred to as Below Poverty Line (BPL) censuses. Households that were automatically considered above the poverty line were not included in these censuses. From among this set, we use the BPL Census 2002 as it is the only dataset, to our knowledge, that provides household-level information on migration patterns.

The fourth such census, the Socioeconomic and Caste Census (SECC), departed from the previous methodology by collecting data on all households, even if they demonstrated characteristics that would exclude them from eligibility under various government schemes

targeted at the poor.[3]

The Government of India has made the SECC publicly available on the internet in PDF and Excel formats. In order to construct a useful microdataset, we scraped over two million files, parsed the embedded text data, and translated these from twelve different Indian languages into English. At the individual level, these data contain variables describing age, gender, occupation, caste group, disability and marital status. At the household level, these data contain variables describing housing, landholdings, agricultural assets, household assets and sources of income. We are able to match these data to our other datasets at the village level. This dataset is unique in describing the economic conditions of every person and household in rural India, at a spatial resolution unavailable from comparable sample surveys.

II.B Population censuses

Since 1871, the Office of the Registrar General of India (ORGI) has conducted a national population census in the first year of every decade. In this paper, we use data from the last three Population Censuses: 1991, 2001 and 2011. The data is reported at the village level. Apart from general demographic characterstics such as village population, age and gender decomposition, caste group, and literacy, the Population Census also provides rich information on village-level amenities and public goods such as paved roads, electricity, primary and secondary schools, health centers, irrigation, bus and rail connectivity et cetera.

II.C Other data

In addition to the socioeconomic and population censuses, we use cross-sectional data from the 68th Round (2011-12) of the National Sample Survey (Employment/Unemployment), which contains far fewer villages and individuals than our census data, but includes data on earnings, place of work and time use across primary and secondary occupations. Using

[3]It is often referred to as the 2011 SECC, as the initial plan was for the survey to be conducted between June and December 2011. However, various delays meant that the majority of the surveying was conducted in 2012, with urban surveys continuing to undergo verification at the time of writing. We therefore use 2012 as the relevant year for the SECC.

village populations backed out from the sample weights, we match observations from the National Sample Survey to the rest of our village-level data.

We use village and town latitude and longitude obtained from ML Infomap to generate measures of straight line distances from villages to towns and district capitals and highways as a proxy for market access. Highway GIS data come from both OpenStreetMap and the National Highways Authority of India.[4]

II.D Rural public goods

Although a number of public goods are relevant, to provide a parsimonious yet informative picture, we focus on paved roads, primary and secondary schools, health centers, and electrification. We use these variables in the binary form: the variable takes the value 1 if the Population Census records the village as having the public good in that year, and 0 otherwise.

II.E Rural economic outcomes

Once again, there are a large number of economic outcomes that we could employ to study the effect of administrative remoteness and the consequent decline in public goods provision. Our selection of economic outcomes is based on availability in the dataset and precise measurement. From the 2012 SECC, we use the share of households whose highest earning member has average monthly income greater than Rs 5000 and Rs 10,000, and the share of households in the village that report having a solid roof (as a proxy for housing quality).

From the Population Censuses, we use the percentage of the village workforce engaged in nonfarm activities, the percentage of village population that is literate, and the share of agricultural land which is irrigated by any source.

Finally, from the BPL Census 2002, we use the share of households in the village that report a household member as being any type of migrant.

[4]We gratefully acknowledge Ejaz Ghani, Arti Goswami and Bill Kerr for generously sharing the GIS data on the Golden Quadrilateral highway network with us.

II.F Calculating average rural income

To the best of our knowledge, there is no publicly available data on incomes at the village level in India. We attempt to overcome this limitation by imputing average monthly income for each village using data from the SECC and the National Sample Survey. For the highest earning member of each household, the SECC reports whether the individual earns less than Rs 5000 (USD 75), between Rs 5000 - 10,000, or more than Rs 10,000 (USD 150). From the 68th Round (2011-12) of the National Sample Survey, we know the precise monthly income for highest-earning members of a nationally representative set of households. We know, for example, that conditional on earning less than Rs 5000, the average monthly income of highest-earning members is Rs 3076; for an individual earning between Rs 5000 - 10,000, the average monthly income is Rs 6,373; and for individuals earning more than Rs 10,000 per month, the average monthly income is Rs 22,353. We use these numbers - along with the share of households in a village whose highest-earning members earn in each of those wage brackets - to calculate a proxy for average monthly income for each village. This is only a proxy for rural incomes, and therefore we do not rely extensively on this measure while reporting our living standard results.

II.G Distance measures

Our main running variable is the village's distance to its district capital. This is the geodesic or straight-line distance in kilometers from the village to the centroid of its district capital.

We also control for village's straight-line distance to the nearest towns of various population sizes (10,000, 50,000, 100,000 and 500,000), and to the nearest highway. These controls serve as proxies for the village's access to relevant urban markets and trunk infrastructure. While we can use actual road distances as opposed to straight line distances, we believe they add to computational costs without enhancing our understanding in a meaningful way.

II.H Local Population Density

We control for population density in the immediate neighborhood of the village. For each village, we calculate the total population that lives within a 0-3 kilometer radius, 3-6 kilometer radius, and so on until 12-15 kilometer radius. For each of these concentric bands, we calculate population density and control for it in our regressions.

II.I Summary statistics

Table 1 shows summary statistics for the full sample of villages. We divide the sample into two halves based on distance to the district capital.

Column 1 contains average values for all villages. Column 2 contains average values for villages whose distance to the district capital is less than the corresponding distance for the median village, while Column 3 reports average values for villages whose distance to the district capital is more than the distance for the median village.

The average village in our sample is 36 kilometers from its district capital and has a population of 1482 in 2011. However there is substantial variation in these averages depending on whether the village's distance to its district capital is more or less than the median. Villages whose distance to district capital is less than the median ("closer" villages) are, on average, 19 kilometers from the capital and have higher average population in 2011 (1548). Villages whose distance to district capital is more than the median ("remoter" villages) are 56 kilometers away on average and are slightly smaller, with an average of 1417 people in 2011.

As we move from the "closer" subsample to the "remoter" subsample, average monthly income decreases by about Rs 450 (approximately US$10, based on average 2011 exchange rates), the share of households in the village with a solid roof decreases by 9 percentage points, and the share of village workforce engaged in nonfarm activities decreases by 9 percentage points. On average, there are no major differences in access to electricity, paved roads, primary schools or medical centers. Villages that are located closer to their district

capital are also closer to a highway (7 kilometers versus 11 kilometers), closer to city with 2011 population exceeding 500,000 (103 versus 113 kms), and closer to a town with 2011 population exceeding 10,000 (12 kms versus 18 kms). Therefore we control for distance to highways, to small towns and to large cities in our regression specifications.

III Empirical Strategy

It is difficult to isolate the effects of administrative remoteness because district capitals are also often the largest towns in the village's catchment area. Further, several measures of connectivity - such as distance to markets, distance to trunk infrastructure, size of local market et cetera - change with distance to district capital.

Therefore we focus our attention on villages located close to district borders. Across these borders, we expect distance to small and large towns, to trunk infrastructure, and local population density to vary smoothly, whereas distance to own district capital - or the degree of "administrative remoteness" - to change discontinuously. We follow the specification in (Dell, 2010) and (Dell et al., 2015) to specify our spatial regression discontinuity equation:

$$\begin{aligned} \mathrm{y}_{v,d,e,s} &= \beta_1 DistCapital_{v,d} + \beta_2 DistHighway_v + \beta_3 DistTown_v \\ &+ \beta_4 Density_{v,15} + f(\text{Geographic Location}_v) + \delta_d + \eta_{d,e,s} + \epsilon_{v,d,e,s} \end{aligned} \tag{1}$$

where $y_{v,d,e,s}$ is the outcome of interest for village v located in district d, near the border between district d and district e, and close to border segment s, i.e. district e is the next closest district to village v. These outcomes include public goods provision outcomes and economic outcomes, as described in the previous section. We restrict our sample to villages lying within a 3 kilometer bandwidth around the district boundaries. Our results are robust to changing this bandwidth to 6 or 9 kilometers.

DistCapital is the geodesic distance in kilometers from village v to the district capital of

its district d. Variation in this variable as we cross the district boundary between district d and district e is what allows us to identify the effect of "administrative remoteness" on our outcomes of interest.

$DistHighway$ is distance in kilometers from village v to the nearest highway. $DistTown$ is geodesic distance in kilometers from village v to the nearest town with 2011 population exceeding 10,000. In some specifications, we also add distance in kilometers to the nearest towns with 2011 population exceeding 50,000, 100,000 and 500,000. $Density_{v,15}$ is the local population density in persons per square kilometer within a 15 kilometer radius of the village.

$f(\text{Geographic Location}_v)$ is the RD polynomial, which controls for smooth functions of geographic location. In our baseline specification, we use a linear polynomial of latitude and longitude. Our results are robust to the choice of RD polynomial: quadratic or cubic.

δ_d is the district fixed effect. Including this in the regression equation allows us to control for any district-specific factors that may affect rural economic outcomes or the provision of public goods. This could include colonial legacy, historical land revenue systems, historical political systems, or district-specific governance measures. The inclusion of the district fixed effect ensures that the coefficient of interest, β_1, captures the effect of distance from district capital that's over and above any district-linked average governance measure, and that our results are not driven by more "administratively remote" villages being located in worse-administered districts.

$\eta_{d,e,s}$ is the border segment fixed effect for the segment of the district d - district e border that lies closest to village v. This is important since the district borders can be quite long - about 40-50 kilometers or more. Some portions of the border may lie closer to one district capital and some portions closer to the other. Hence it is important to segment the border, to ensure our comparison groups are located within the same small geographical region. We construct border segments by intersecting grid cells of approximately 15 kms per side with the district border. When we restrict our sample to villages lying within 3 kilometers of the

district boundary, this is approximately equal to identifying variation in distance to district capital within a 6 km x 15 km grid-cell.

Following (Burgess et al., n.d.), we cluster the errors in blocks of size 50km by 50km to allow for some geographical error correlation.

Figure 1 best illustrates our empirical strategy. The figure is a map of a few districts in Rajasthan: Alwar, Jaipur, Dausa, Sikar and Karauli, in which their district capitals have been represented by black dots. Different colored dots represent villages located near different segments of the district border. These are all villages located within 3 kilometers of the district border. Filled circles are villages located on the side of the district border which is closer to the district capital, and hollow circles are villages located on the side of the district border that's more distant from the district capital. We can observe, for example, that along the Jaipur-Dausa border, villages on the Dausa side are closer to their district capital than villages on the Jaipur side of the border. Along the Jaipur-Alwar border, the southernmost villages are located closer to Jaipur's district capital, whereas as we move north, the Alwar side of the border becomes more proximate to the district capital. Our variation comes from a change in distance to district capital that occurs within the border segment, i.e. among villages with the same color in this Figure.

III.A Balance

We want to see how the main control variables in Equation 1, such as distance to nearest town with population above certain thresholds, and balance variables such as percentage of village population that belongs to Scheduled Castes (SC), change as we cross the district boundary. Large changes in these variables would violate our assumption that the effects on rural outcomes are driven by distance to centers of administration alone, and not by distance to urban markets in general.

To test this, we construct a binary version of the distance to district capital variable. For

each border segment, we calculate average distance from villages on both sides of the border to the corresponding district capital. We consider the side with the smaller average distance to district capital as the "closer side", and estimate the following regression equation:

$$\mathrm{y}_{v,d,e,s} = \beta_1 CloserSide + f(\text{Geographic Location}_v) + \delta_d + \epsilon_{v,d,e,s} \tag{2}$$

where $y_{v,d,e,s}$ is the control for village v located in district d, near the border between district d and district e, and close to border segment s. These controls include distance to nearest town with 2011 population exceeding 10,000, 50,000, 100,000 and 500,000, as well as percent SC in the village population.

$f(\text{Geographic Location}_v)$ is the RD polynomial as before, and δ_d is the district fixed effect. Standard errors have been clustered the same way as in the main regression.

IV Results

In this section, we describe and discuss the countrywide spatial gradients in rural outcomes (Section IV.A), the balance checks (Section IV.B), the main results (Section IV.D), robustness (Section IV.D) and the evidence on the mechanism (Section IV.E). We first show that rural outcomes worsen fairly monotonically with distance to towns. We then focus on our spatial regression discontinuity specification, starting with the balance checks. In the main results, we show that distance from district capital reduces public goods provision in villages and worsens socioeconomc outcomes, by comparing villages located in close proximity on either side of a district boundary. We then show that these results are not driven by differences between villages on either side of a state boundary, the bandwidth around the district boundary, or the manner in which we the geographical location of the villages enters the regression specification. We then consider an important mechanism that could explain these results, finding that at this stage, the evidence best supports higher cost of constructing

public assets in these villages.

IV.A Spatial gradients

Figures 2 and 3 show the binscatters for a range of rural outcomes as a function of distance to nearest town with 2011 population exceeding 10,000 and 100,000. We can observe monotonically negative spatial gradients for average rural earnings, proportion of households in the village with solid roof, percentage of village population that is literate, probability that the village is electrified, share of total land which is irrigated by any source, and probability that the village has a paved road. The gradients are flatter for distance to towns with 2011 population exceeding 100,000, but show a remarkably similar pattern. There seems to be a substantial cost to villages from greater distance to urban markets, and this pattern can be systematically observed for the entire country.

IV.B Balance checks

Table 2 shows estimates from Equation 2 where we regress various distance controls on the binary variable $CloserSide$, which takes the value 1 for villages that belong to the side of the border segment that is on average closer to its district capital, and 0 otherwise. The first column confirms that we have correctly identified the closer side of the border segment: going from the not-close side to the closer side brings the villages on average 17.8 kilometers closer to their district capital. The other columns show that while more administratively proximate villages are, on average, also closer to small and large towns, the difference is only a few hundred meters. For example, administratively proximate villages are 291 meters closer to towns with 2011 population exceeding 10,000, which is about 1.7% of the average distance to towns with 2011 population exceeding 10,000 in our sample. The corresponding percentage for towns with 2011 population exceeding 50,000 is 2.2%, and for towns with 2011 population exceeding 100,000 is 1.4%. We also observe that proportion of village population that belongs to Scheduled Castes is balanced across the more administratively proximate

and administratively remote sides of the border segments.

IV.C Main results

Table 3 presents estimates from Equation 1 for the effect of distance from district capital on rural public goods provision. We note that while the probability of receiving electricity and having schools or a medical center is not systematically different between more and less administratively proximate villages, villages face considerable cost of administrative remoteness in terms of reduced access to paved roads. Since the standard deviation of distance to district capital is approximately 23.86 kilometers, a one standard deviation increase in administrative remoteness reduces the probability that the average village has a paved road by 1.9 percentage points.

Table 4 presents estimates from Equation 1 for the effect of distance from district capital on rural economic outcomes. A one standard deviation increase in distance to district capital reduces the proportion of workforce in nonfarm activities by 3.1 percentage points and average literacy rates by 1 percentage point. Crucially, distance to district capital does not matter for economic outcomes such as mean monthly earnings and housing quality. We expect these economic outcomes to be correlated negatively with distance to towns, which they are, but not with access to the state. This supports our hypothesis that our results are driven by distance to administrative centers and not urban centers in general.

IV.D Robustness

In this section, we show the robustness of our results to omitting those district borders that are also state borders, to changing the specification of the geographic location polynomial, and to changing the bandwidth around the district boundary.

Several state borders in India are also linguistic or ethnic borders, and therefore we may be concerned that the results are driven by state-specific unobservable factors that are not picked up by the district fixed effects. Mobility across state borders is also more limited than

mobility across district borders within the same state (Internal Borders and Migration in India, . Hence it is important to repeat the estimation of Equation 1 for the effect of distance from district capital on rural public goods and economic outcomes, for only those district borders that lie completely within the state. We lose some power by omitting inter-state district borders, but our main results: the negative effects of distance to district capital on paved roads, nonfarm employment and literacy rates, still hold, as can be seen from Tables 6 and 7.

Further, we may be concerned that the results seen in the may be contingent on the specific form of the geographic polynomial we chose: a linear polynomial in latitude and longitude. In Tables 8 and 9 we observe the same results as before using a quadratic polynomial in latitude and longitude. Tables 10 and 11 estimate Equation 1 using a cubic polynomial in latitude and longitude, and once again, distance to district capital reduces the probability of paved roads in the village, the share of the village's workforce engaged in nonfarm activities, and its literacy rate.

Finally, the results are robust to changing the size of the bandwidth around the district border from 3 km to 6 or 9 km. This is demonstrated by Figure 4, which contains a coefficient plot for the main variables of interest. We observe that our coefficients or their statistical significance does not change meaningfully with a change in bandwidth from 3 to 6 or 9 kilometers. The consistency of coefficient magnitudes also suggests that these effects are not driven by variation in distance to district capital within the same district.

IV.E Mechanism

The results we observe can be explained by several mechanisms. Villages located at greater distances from the district capitals may be receiving fewer public goods because of the high cost of providing district-specific services in these places. For example, if these places have bad roads to begin with, the cost of paving roads or building new ones will be correspondingly higher. Another mechanism is the higher cost of monitoring government programs in these

places. District public servants may find it less costly to visit villages within a day's travelling from the district capital, but may visit farther places less frequently. This can reduce the provision of public services in more administratively remote locations. Finally, it may be the citizens living in more administratively remote locations are less informed about government policies and this can reduce their ability to organize and demand public goods such as paved roads. Krishna and Schober (2014) finds evidence for such a mechanism in southern India.

At this stage, we provide evidence for the first mechanism - the higher cost of providing public goods in more administratively remote locations. These costs are usually hard to observe for districts across the country, and even harder to compare across districts. However, for one public good - paved roads - we can exploit project data from the PMGSY project to say something about how unit costs change as a function of distance to district capital.

We assemble data on cost and duration of construction and road length in kilometers from over 100,000 roads constructed under the PMGSY project. We use this to calculate per kilometer measures of road cost and construction duration. Table 5 reports estimates from regressing PMGSY variables on distance to district capital for villages located close to district borders. We find that a one standard deviation increase in distance to district capital increases the cost per kilometer by 2.2%. This provides evidence for the first channel through which administrative remoteness may affect rural public goods provision in India.

V Conclusion

Citizens in developing countries have unequal access to public goods and services, and this inequality varies systemtically across space. The structure of governance, which determines how public goods are provided, contributes to this inequality.

In this paper, we estimate the cost to the rural economy of being located at a greater distance from the local administrative center, or "administrative remoteness". To do this, we assemble a rich panel dataset on rural public goods, household economic outcomes, de-

mographic characteristics and geographical location for all villages in India. We isolate the effects of administrative remoteness by focussing on border areas of districts, which are responsible for implementing most public programs in India. Access to urban markets, trunk infrastructure and population density vary smoothly across the district border, but distance to district capital varies sharply. We use this geographical discontinuity to isolate the causal effects of administrative remoteness on rural outcomes.

We find that administrative remoteness has a negative effect on the provision of public goods and economic outcomes in rural India. Villages located at greater distances from their district capital have a lower probability of receiving a paved road or a secondary school compared to neighboring villages that are located substantially closer to their district capital. Villages that are more administratively remote also have significantly lower average income, smaller share of households with a solid roof, lower literacy rates and a lower percentage of the workforce engaged in nonfarm activities. We find these results to be robust to a range of alternative specifications. Evidence from the PMGSY rural roads project suggests that one mechanism driving these effects is the higher cost of building public infrastructure - such as paved roads - in villages that are located at greater distances from their district capital. Further work remains to be done to uncover the factors driving this higher cost in more administratively remote locations, as well as other mechanisms through which the cost of remoteness operates.

Put together, our results suggest that public administration plays an important role in contributing to the spatial inequality in access to public goods - even within fairly narrow geographical areas - and this has a negative effect on rural economic outcomes. Further work needs to be done to recommend policy reforms that can reduce this spatial inequality in developing countries.

Table 1: Summary statistics

	Full Sample	Low Distance to District HQ	High Distance to District HQ
Mean monthly earnings (2012 Rupees)	5076.2	5298.8	4853.1
	(2400.4)	(2495.0)	(2280.0)
Percent households with solid roof (2012)	48.87	53.17	44.56
	(34.17)	(33.27)	(34.52)
Paved Road Access (2011)	80.11	81.52	78.69
	(39.92)	(38.82)	(40.95)
Percent population literate (2011)	57.79	59.14	56.45
	(13.91)	(13.64)	(14.04)
Percent workforce in nonfarm activities (2011)	28.92	33.27	24.57
	(27.04)	(28.08)	(25.22)
Percent land irrigated (2011)	58.18	61.41	54.95
	(38.51)	(38.59)	(38.16)
Percent villages electrified (2011)	59.48	58.19	60.77
	(49.09)	(49.32)	(48.83)
Percent villages with govt secondary school (2011)	16.16	15.58	16.75
	(36.81)	(36.27)	(37.34)
Population (2011)	1482.1	1547.6	1416.5
	(1996.2)	(2114.9)	(1867.5)
Distance to nearest town (kms)	15.39	12.72	18.06
	(10.26)	(6.825)	(12.24)
Distance to nearest town with pop over 500k (kms)	108.0	103.3	112.8
	(63.74)	(66.86)	(60.07)
Distance to District HQ (kms)	36.44	19.25	53.67
	(22.78)	(7.675)	(19.66)
Observations	440104	220278	219826

Notes: This table presents means and standard deviations for observed outcomes for all villages in our sample. The 2002 data is from the BPL Census 2002, 2011 data from the Population Census 2011, and the 2012 data from the Socioeconomic Census 2012. The "closer villages" column presents values for villages whose distance to their district capital is less than distance to district capital for the median village. The "remoter villages" column presents values for villages whose distance to their district capital is larger than distance to district capital for the median village.

Table 2: Balance checks

	Distance to DHQ	Distance 10k town	Distance 50k town	Distance 100k town	Distance to Highway	Percent SC
Closer side of border	-17.769	-0.291	-0.730	-0.672	-0.217	-0.281
	(0.512)***	(0.062)***	(0.067)***	(0.067)***	(0.054)***	(0.204)
Outcome Mean	42.63	16.21	33.13	48.75	9.691	18.69
Bandwidth	3 km	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear	Linear	Linear
N	100659	100659	100659	100659	100659	100659
R2	.9364	.9381	.9797	.9915	.9173	.3885

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

Notes: This table presents regression estimates from regressing a binary variable describing whether or not the village is on the closer to district capital side of a district border segment, on distance to district capital and a range of variables used as controls in Equation **??**. All regressions include district fixed effects and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 3: Rural public goods provision

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.063	0.005	-0.020	-0.016	0.004
	(0.021)***	(0.027)	(0.014)	(0.014)	(0.018)
Distance to nearest town (kms)	0.013	-0.015	0.038	0.106	0.014
	(0.061)	(0.077)	(0.052)	(0.045)**	(0.055)
Distance to nearest highway (kms)	-0.238	-0.137	-0.024	-0.045	-0.186
	(0.065)***	(0.091)	(0.049)	(0.050)	(0.058)***
Outcome Mean	80.69	55.75	85.95	14.04	18.81
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear	Linear
Density controls	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100659	100659
R2	.4125	.6302	.268	.2018	.2674

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

Notes: The table presents regression estimates from Equation 1, where we regress public goods provision on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 4: Rural economic outcomes

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Migrant households
Distance to District HQ (kms)	-1.396	-0.005	-0.022	-0.036	-0.024	0.015
	(1.096)	(0.012)	(0.005)***	(0.010)***	(.)	(0.021)
Distance to nearest town (kms)	-5.941	-0.127	-0.079	-0.246	-0.107	0.056
	(3.143)*	(0.036)***	(0.019)***	(0.044)***	(.)	(0.059)
Distance to nearest highway (kms)	-10.969	-0.275	-0.153	-0.342	-0.085	0.159
	(3.400)***	(0.040)***	(0.020)***	(0.047)***	(.)	(0.062)**
Outcome Mean	4986	47.63	56.93	27.83	58.67	61.6
Bandwidth	3 km	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear	Linear	Linear
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100312	97719	70111
R2	.4058	.7587	.671	.4761	.7673	.5627

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

Notes: The table presents regression estimates from Equation 1, where we regress rural economic outcomes on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Mean income refers to imputed average monthly income based on assigning monthly income of Rs 3,076 to households whose highest earning member reports monthly income of less than Rs 5,000 in the 2012 SECC, Rs 6,373 to households whose highest earning member reports monthly income greater than Rs 5,000 but less than Rs 10,000 in the 2012 SECC, and Rs 22,353 to households whose highest earning member reports monthly income greater than Rs 10,000 in the 2012 SECC. These precise numbers are conditional monthly income averages for earners in these wage ranges as reported by the 68th Round (2011-12) of the National Sample Survey. Solid roof refers to share of households in the village that report having a solid roof in the 2012 SECC. Percent Literate refers to the village population classified as literate in the 2011 Population Census. Percent Nonfarm refers to the proportion of village main workers that are engaged in nonfarm activities as reported by the 2011 Population Census. Percent Land Irrigated is the share of village agricultural land that is irrigated as per the 2011 Population Census. Households with a migrant is the share of households in the village that report at least one family member as a migrant in the 2002 BPL Census. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 5: Road construction costs

	Cost per km	Cost Overrun per km	Time Overrun per km	Time per km
Distance to District HQ (kms)	0.003	-0.001	0.152	0.220
	(0.001)**	(0.001)*	(0.226)	(0.286)
Distance to nearest town (kms)	-0.006	0.000	1.074	0.334
	(0.005)	(0.003)	(0.891)	(1.098)
Distance to nearest highway (kms)	0.009	-0.001	0.080	-0.878
	(0.005)*	(0.003)	(0.966)	(1.134)
Outcome Mean	3.22	-.2038	88.01	234.1
Bandwidth	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear
Density controls	Yes	Yes	Yes	Yes
N	17321	13079	14616	14593
R2	.7349	.5332	.4949	.5091

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

Notes: The table presents regression estimates from Equation 1, where we regress PMGSY program outcomes on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Cost per kilometer is the final cost of constructing the PMGSY road in million rupees divided by the length of the road in kilometers. Cost overrun per kilometer is the difference between the estimated cost and the projected cost divided by the length of the road in kilometers. Time overrun per kilometer is the difference between actual completion date and projected completion date divided by the length of the road in kilometers. Time per kilometer is the difference between actual completion date and project start date divided by length of the road in kilometers. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 6: Rural public goods provision, omitting state borders

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.100	0.024	-0.006	-0.005	0.014
	(0.027)***	(0.028)	(0.017)	(0.017)	(0.021)
Distance to nearest town (kms)	0.026	-0.011	0.066	0.119	0.002
	(0.067)	(0.083)	(0.059)	(0.051)**	(0.059)
Distance to nearest highway (kms)	-0.239	-0.131	-0.043	-0.018	-0.184
	(0.075)***	(0.104)	(0.056)	(0.056)	(0.063)***
Outcome Mean	81.31	56.83	85.84	14.41	18.65
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear	Linear
Density controls	Yes	Yes	Yes	Yes	Yes
N	81531	81531	81531	81531	81531
R2	.4042	.6212	.2682	.2034	.2712

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

Notes: The table presents regression estimates from Equation 1, where we regress public goods provision on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. District borders that are also state borders are omitted from this regression. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 7: Rural economic outcomes, omitting state borders

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Migrant households
Distance to District HQ (kms)	-2.134	-0.001	-0.016	-0.033	-0.044	0.010
	(1.219)*	(0.014)	(0.006)***	(0.012)***	(0.020)**	(0.024)
Distance to nearest town (kms)	-4.792	-0.138	-0.084	-0.221	-0.116	0.049
	(3.500)	(0.042)***	(0.021)***	(0.049)***	(0.058)**	(0.065)
Distance to nearest highway (kms)	-12.397	-0.267	-0.133	-0.348	-0.044	0.147
	(3.650)***	(0.044)***	(0.023)***	(0.054)***	(0.067)	(0.070)**
Outcome Mean	5017	48.54	58	28.17	60.27	61.71
Bandwidth	3 km	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear	Linear	Linear
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	81531	81531	81531	81276	79239	57314
R2	.4078	.7486	.6655	.4721	.7711	.5538

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

Notes: The table presents regression estimates from Equation 1, where we regress rural economic outcomes on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. District borders that are also state borders are omitted from this regression. Mean income refers to imputed average monthly income based on assigning monthly income of Rs 3,076 to households whose highest earning member reports monthly income of less than Rs 5,000 in the 2012 SECC, Rs 6,373 to households whose highest earning member reports monthly income greater than Rs 5,000 but less than Rs 10,000 in the 2012 SECC, and Rs 22,353 to households whose highest earning member reports monthly income greater than Rs 10,000 in the 2012 SECC. These precise numbers are conditional monthly income averages for earners in these wage ranges as reported by the 68th Round (2011-12) of the National Sample Survey. Solid roof refers to share of households in the village that report having a solid roof in the 2012 SECC. Percent Literate refers to the village population classified as literate in the 2011 Population Census. Percent Nonfarm refers to the proportion of village main workers that are engaged in nonfarm activities as reported by the 2011 Population Census. Percent Land Irrigated is the share of village agricultural land that is irrigated as per the 2011 Population Census. Households with a migrant is the share of households in the village that report at least one family member as a migrant in the 2002 BPL Census. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 8: Rural public goods provision using quadratic polynomial

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.063	0.005	-0.020	-0.016	0.004
	(0.021)***	(0.026)	(0.014)	(0.014)	(0.018)
Distance to nearest town (kms)	0.012	-0.015	0.039	0.105	0.013
	(0.061)	(0.077)	(0.052)	(0.045)**	(0.055)
Distance to nearest highway (kms)	-0.239	-0.136	-0.024	-0.045	-0.186
	(0.064)***	(0.091)	(0.050)	(0.049)	(0.059)***
Outcome Mean	80.69	55.75	85.95	14.04	18.81
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Density controls	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100659	100659
R2	.4125	.6302	.268	.2018	.2674

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

Notes: The table presents regression estimates from Equation 1, where we regress public goods provision on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a quadratic polynomial in latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 9: Rural economic outcomes using quadratic polynomial

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Migrant households
Distance to District HQ (kms)	-1.389	-0.005	-0.022	-0.036	-0.024	0.015
	(1.096)	(0.012)	(0.005)***	(0.010)***	(.)	(0.021)
Distance to nearest town (kms)	-5.934	-0.128	-0.080	-0.247	-0.106	0.056
	(3.143)*	(0.036)***	(0.019)***	(0.043)***	(.)	(0.059)
Distance to nearest highway (kms)	-10.952	-0.276	-0.153	-0.343	-0.084	0.158
	(3.393)***	(0.040)***	(0.020)***	(0.047)***	(.)	(0.062)**
Outcome Mean	4986	47.63	56.93	27.83	58.67	61.6
Bandwidth	3 km	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100312	97719	70111
R2	.4058	.7588	.6711	.4762	.7673	.5627

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

Notes: The table presents regression estimates from Equation 1, where we regress rural economic outcomes on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a quadratic polynomial in latitude and longitude. Mean income refers to imputed average monthly income based on assigning monthly income of Rs 3,076 to households whose highest earning member reports monthly income of less than Rs 5,000 in the 2012 SECC, Rs 6,373 to households whose highest earning member reports monthly income greater than Rs 5,000 but less than Rs 10,000 in the 2012 SECC, and Rs 22,353 to households whose highest earning member reports monthly income greater than Rs 10,000 in the 2012 SECC. These precise numbers are conditional monthly income averages for earners in these wage ranges as reported by the 68th Round (2011-12) of the National Sample Survey. Solid roof refers to share of households in the village that report having a solid roof in the 2012 SECC. Percent Literate refers to the village population classified as literate in the 2011 Population Census. Percent Nonfarm refers to the proportion of village main workers that are engaged in nonfarm activities as reported by the 2011 Population Census. Percent Land Irrigated is the share of village agricultural land that is irrigated as per the 2011 Population Census. Households with a migrant is the share of households in the village that report at least one family member as a migrant in the 2002 BPL Census. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Table 10: Rural public goods provision using cubic polynomial

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.064	0.005	-0.020	-0.016	0.003
	(0.021)***	(0.026)	(0.014)	(0.014)	(0.018)
Distance to nearest town (kms)	0.015	-0.013	0.038	0.106	0.014
	(0.061)	(0.077)	(0.052)	(0.045)**	(0.055)
Distance to nearest highway (kms)	-0.239	-0.136	-0.024	-0.045	-0.186
	(0.064)***	(0.091)	(0.050)	(0.049)	(0.059)***
Outcome Mean	80.69	55.75	85.95	14.04	18.81
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Cubic	Cubic	Cubic	Cubic	Cubic
Density controls	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100659	100659
R2	.4126	.6302	.2681	.2018	.2675

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

Notes: The table presents regression estimates from Equation 1, where we regress public goods provision on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a cubic polynomial in latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

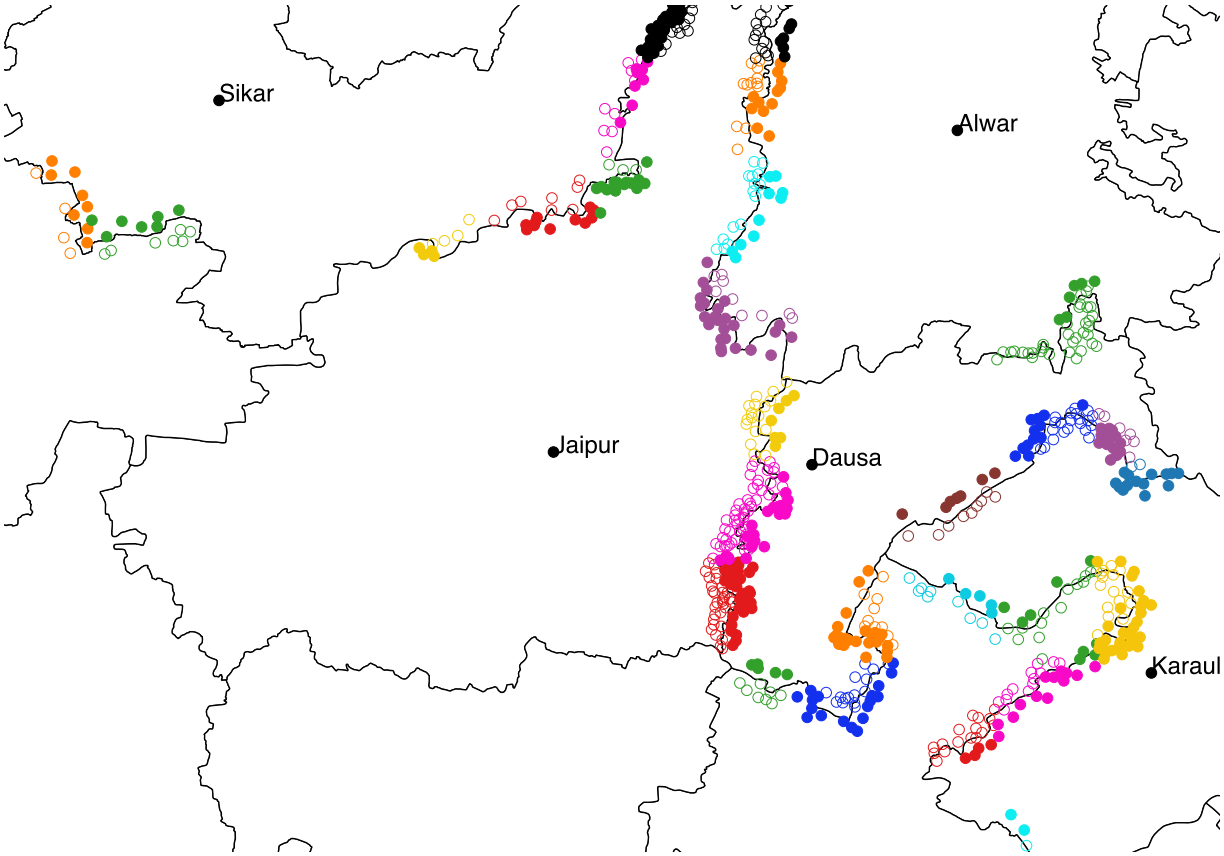
Table 11: Rural economic outcomes using cubic polynomial

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Migrant households
Distance to District HQ (kms)	-1.410	-0.005	-0.022	-0.036	-0.023	0.015
	(1.094)	(0.012)	(0.005)***	(0.010)***	(.)	(0.021)
Distance to nearest town (kms)	-5.766	-0.128	-0.079	-0.247	-0.108	0.054
	(3.146)*	(0.036)***	(0.019)***	(0.043)***	(.)	(0.059)
Distance to nearest highway (kms)	-10.957	-0.276	-0.153	-0.343	-0.084	0.158
	(3.386)***	(0.040)***	(0.020)***	(0.047)***	(.)	(0.062)**
Outcome Mean	4986	47.63	56.93	27.83	58.67	61.6
Bandwidth	3 km	3 km	3 km	3 km	3 km	3 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	100659	100659	100659	100312	97719	70111
R2	.4059	.7588	.6711	.4762	.7673	.5627

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

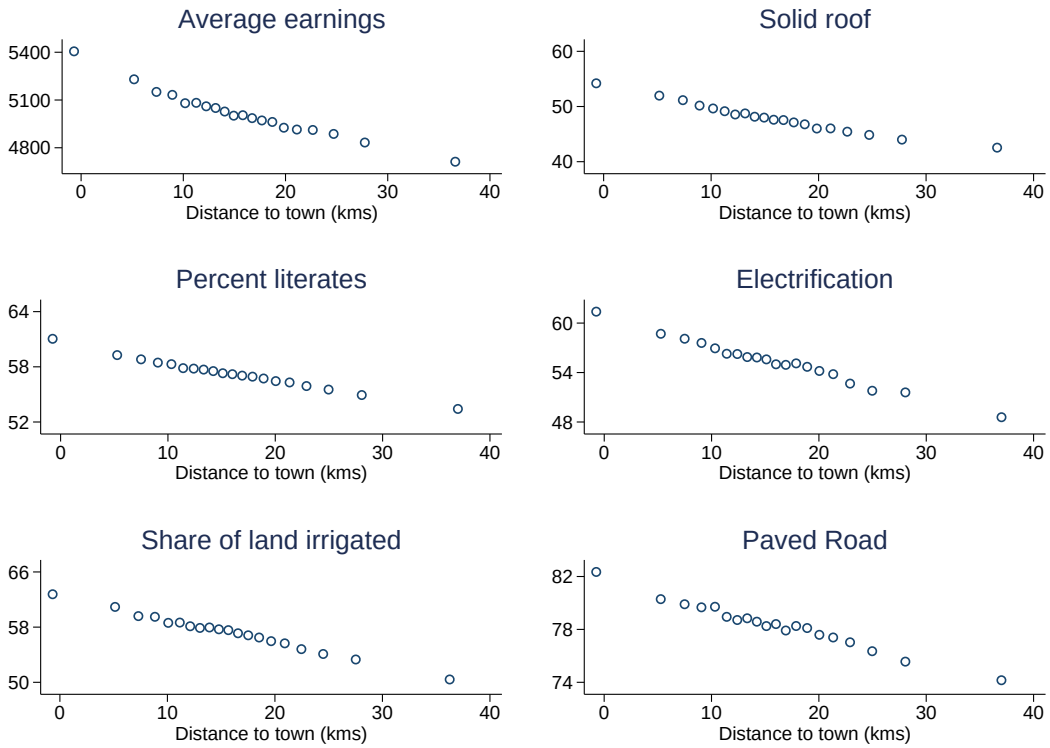
Notes: The table presents regression estimates from Equation 1, where we regress rural economic outcomes on distance to district capital in kilometers, distance to nearest town, distance to highways, local population density, and include district and border segment fixed effects. We control for geographic location using a cubic polynomial in latitude and longitude. Mean income refers to imputed average monthly income based on assigning monthly income of Rs 3,076 to households whose highest earning member reports monthly income of less than Rs 5,000 in the 2012 SECC, Rs 6,373 to households whose highest earning member reports monthly income greater than Rs 5,000 but less than Rs 10,000 in the 2012 SECC, and Rs 22,353 to households whose highest earning member reports monthly income greater than Rs 10,000 in the 2012 SECC. These precise numbers are conditional monthly income averages for earners in these wage ranges as reported by the 68th Round (2011-12) of the National Sample Survey. Solid roof refers to share of households in the village that report having a solid roof in the 2012 SECC. Percent Literate refers to the village population classified as literate in the 2011 Population Census. Percent Nonfarm refers to the proportion of village main workers that are engaged in nonfarm activities as reported by the 2011 Population Census. Percent Land Irrigated is the share of village agricultural land that is irrigated as per the 2011 Population Census. Households with a migrant is the share of households in the village that report at least one family member as a migrant in the 2002 BPL Census. Standard errors are clustered in blocks of size 50km by 50km to allow for some geographical error correlation.

Figure 1: Illustration of Empirical Strategy



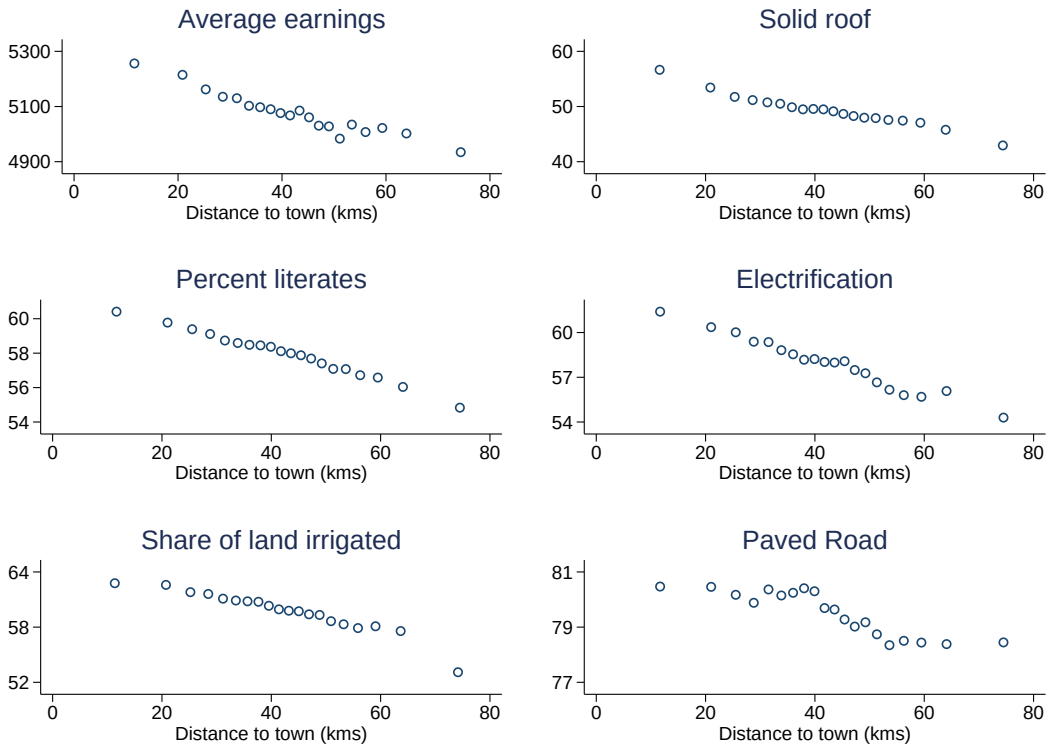
Notes: Example of our empirical strategy, showing different districts with their district capitals and border villages in some districts in Rajasthan. Different colored dots represent different segments of the district border. Filled circles are villages located on the proximate side of the district border, hollow circles are villages located on the more remote side of the district border.

Figure 2: Costs of remoteness in India



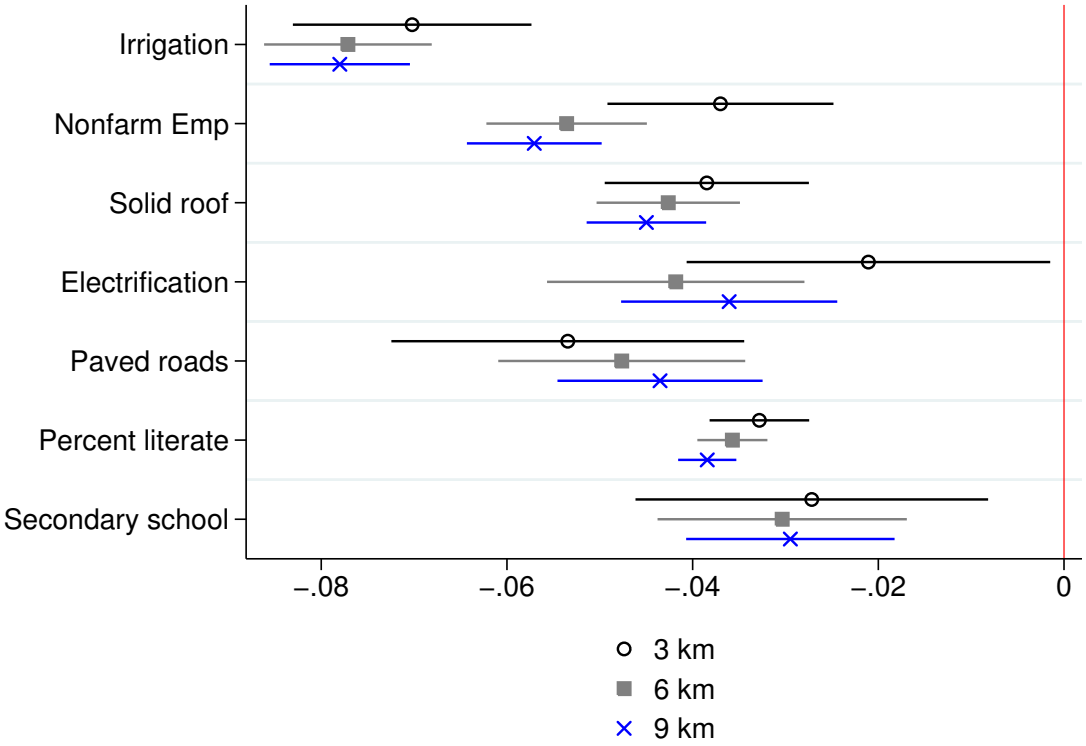
Notes: These graphs show the monotonic decline in a range of rural outcomes with distance from towns with population exceeding 10,000

Figure 3: Costs of remoteness in India



Notes: These graphs show the monotonic decline in a range of rural outcomes with distance from towns with population exceeding 100,000

Figure 4: Coefficient Plot for different bandwidths



Notes: This coefficient plot shows that our results are robust to changing the bandwidth around the district boundary. The similarity of the magnitudes suggests that our results are not driven by changes in distance to district capital within the same district.

References

Alkire, Sabina and Suman Seth, "Identifying BPL Households: A Comparison of Methods," 2012.

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

Bardhan, Pranab, "Decentralization of governance and development," *The Journal of Economic Perspectives*, 2002, *16* (4), 185–205.

Bates, Robert H, "Modernization, ethnic competition, and the rationality of politics in contemporary Africa," *State versus ethnic claims: African policy dilemmas*, 1983, *152*, 171.

Bryan, Gharad and Melanie Morten, "Economic development and the spatial allocation of labor: Evidence from Indonesia," *Manuscript, London School of Economics and Stanford University*, 2015.

Burgess, Robin, Francisco Costa, and Benjamin Olken, "The Power of the State: National Borders and the Deforestation of the Amazon."

Campante, Filipe R et al., "Isolated capital cities, accountability, and corruption: Evidence from US states," *The American Economic Review*, 2014, *104* (8), 2456–2481.

Dell, Melissa, "The persistent effects of Peru's mining mita," *Econometrica*, 2010, *78* (6), 1863–1903.

____, Nathan Lane, and Pablo Querubin, "State Capacity, Local Governance, and Economic Development in Vietnam," *NBER Working Paper, pp. l-40*, 2015.

Fafchamps, Marcel and Forhad Shilpi, "The spatial division of labour in Nepal," *The Journal of Development Studies*, 2003, *39* (6), 23–66.

Feyrer, James, "Distance, tradet, and income-the 1967 to 1975 closing of the Suez Canal as a natural experiment," Technical Report, National Bureau of Economic Research 2009.

Herbst, Jeffrey, *States and power in Africa: Comparative lessons in authority and control*, Princeton University Press, 2014.

Hnatkovska, Viktoria and Amartya Lahiri, "Structural transformation and the rural-urban divide," *University of British Columbia, typescript*, 2013.

Internal Borders and Migration in India.

Jalan, Jyotsna, Martin Ravallion et al., *Spatial poverty traps*, Citeseer, 1997.

Kanbur, Ravi and Anthony J Venables, *Spatial inequality and development*, OUP Oxford, 2005.

____ and Hillel Rapoport, "Migration selectivity and the evolution of spatial inequality," *Journal of Economic Geography*, 2005, *5* (1), 43–57.

Krishna, Anirudh and Gregory Schober, "The gradient of governance: distance and disengagement in Indian villages," *Journal of Development Studies*, 2014, *50* (6), 820–838.

Michalopoulos, Stelios and Elias Papaioannou, "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, 2014, *129* (1), 151–213.

Moretti, Enrico, "Local labor markets," *Handbook of labor economics*, 2011, *4*, 1237–1313.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," *The Review of Economic Studies*, 2016, p. rdw020.

The World Bank, "World Development Report 2009: Reshaping Economic Geography," Technical Report, World Bank 2009.

——— , "The Rural Access Index," Technical Report, World Bank 2015.

Young, Alwyn, "Inequality, the urban-rural gap and migration," *The Quarterly Journal of Economics*, 2013, p. qjt025.