The Cost of Distance: Geography and Governance in Rural India* PRELIMINARY: PLEASE DO NOT CITE WITHOUT PERMISSION

Sam Asher† Karan Nagpal‡ Paul Novosad§

October 28, 2016

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 headquarters. We exploit administrative boundaries that generate sharp jumps in distance to administrative headquarters but not in market access, population density or distance to trunk infrastructure. Villages that are more distant from administrative headquarters receive fewer paved roads and secondary schools, have lower literacy and more limited participation in non-agricultural activities. These effects are driven by the higher cost of building infrastructure, such as roads, in more distant villages.

JEL Codes: R12, D63, H41, O18

*This version October 2016. We thanks seminar participants at the Center for Global Development, the DFID-IZA Workshop in Oxford, and at the Department for International Development in 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

1 Introduction

Economic outcomes vary systematically across space (Bryan and Morten, 2015; Moretti, 2011; Kanbur and Venables, 2005; Kanbur and Rapoport, 2005). Spatial inequality is particularly intense in developing countries, where average household consumption in richer regions can be almost 75% higher than in poorer regions of the same country; the corresponding differential for developed countries is less than 25% (Bank, 2009). In addition to income and consumption, there is also substantial 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 (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 evidence that the distance of a village from its administrative headquarter, 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, with information on public goods, average earnings, household assets, employment structure and geographic location over a 21-year period (1991-2012). We calculate each village's distance to its district headquarters. This matters for public goods provision because most public programs in India are implemented by the district administration which is based in the district headquarters. There were 640 districts in India in 2011, with an average of approximately two million citizens per district. These districts are shown in Figure 1.

The probability of receiving various public goods, such as paved roads, electricity, primary and secondary schools, and health centers is negatively correlated with distance to district headquarters. This distance is also negatively correlated with rural economic outcomes such as average income, housing quality, literacy, and the percentage of village workforce engaged in non-agricultural activities.

We implement a spatial regression discontinuity design to estimate the causal effects of distance from district headquarters. We compare villages on either side of district borders, such that distance to urban markets, distance to trunk infrastructure and local population density vary smoothly across the border. However, there is a sharp jump in distance to district headquarter, due to variation in the geographical location of the district headquarter within each district.

We find that an increase in the distance to district headquarter reduces the provision of public goods that are managed by the district administration, such as paved roads and secondary schools. It does not affect public goods provided by a higher tier of administration, such as electrification. Increase in distance to district headquarters also affects economic outcomes adversely, causing a reduction in average rural income, housing quality, literacy rate and the proportion of rural workforce engaged in non-agricultural activities. For example, a one standard deviation increase in distance to district headquarters (a change of about 24 kilometers) reduces the probability of paved road connections by 1.4%, and probability of secondary school by 6.2%. It also reduces literacy by 0.8%, proportion of workforce in nonfarm activities by 1.4%, and share of households in the village with a solid roof by 1.4%. Our results are robust to changing the distance bandwidth around the district borders, as well as to the regression discontinuity specification.

The reduced provision of public goods is linked to higher unit costs of provision. We assemble data on cost per kilometer and duration of construction for rural roads under the largest rural roads construction program in India, the Pradhan Mantri Gram Sadak Yojna (PMGSY). Preliminary results suggest that a one standard deviation increase in distance to district headquarters increases the cost per kilometer under the PMGSY program by 1.54%.

This paper contributes 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.

Our paper also adds to a literature that studies 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 capacity" 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 mnay developing countries.

There is also a long-standing literature in economics on the costs and benefits from decentralization (Bardhan, 2002). In this paper, we show that distance from district headquarter matters for public goods provided by the district administration, such as paved roads, but not for public goods provided by higher tiers of administration, such as electricity, which is provided by a federal agency (Rural Electrification Corporation).

The rest of the paper proceeds as follows: Section 2 describes data construction and main variables of interest. Section 3 explains the empirical strategy. Section 4 presents and discussed our results, including robustness checks. Section 5 concludes.

2 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). To do this, 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.

2.1 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.[1]

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

[1] 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.

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.

2.2 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.

2.3 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 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 headquarters and highways as a proxy for market access. Highway GIS data come from both OpenStreetMap and the National Highways Authority of India.[2]

2.4 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. In some specifications, we also use road quality from PMGSY administrative data and number of hours of electricity from the 2011 Population Census (previous censuses have not recorded this variable).

2.5 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.

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

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

2.6 Calculating average rural income

To the best of our knowledge, there is no publicly available data on incomes at the village level in India, or indeed any other large developing country. 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.

2.7 Distance measures

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

We also control for village's straight-line distance to the nearest town with population greater than 10,000 in 2011, 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.

2.8 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.

2.9 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 headquarters.

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

The average village in our sample is 38 kilometers from its district headquarters and has a population of 1,485 people in 2011. However there is substantial variation in these averages depending on whether the village's distance to its district headquarters is more or less than the median. Villages whose distance to district headquarters is less than the median ("closer" villages) are, on average, 20 kilometers from the headquarters and have higher average population in 20111 (1579). Villages whose distance to district headquarters is more than the median ("remoter" villages) are 56 kilometers away on average and are slightly smaller, with an average of 1,406 people in 2011.

As we move from the "closer" subsample to the "remoter" subsample, average monthly income decreases by about Rs 400, the share of households in the village with a solid roof decreases by 10 percentage points, and the share of village workforce engaged in nonfarm activities decreases by 7 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 headquarters are also closer to a highway (7 kilometers versus 11 kilometers). Therefore we control for access to trunk infrastructure in our regression specifications.

3 Empirical Strategy

It is difficult to isolate the effects of administrative remoteness because district headquarters can also often be 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 headquarters.

Therefore we focus our attention on villages located close to district borders. Access to markets, access to trunk infrastructure, and local population density vary smoothly across a district border, whereas there is a discontinuous jump in distance to the relevant district headquarter, or the degree of administrative remoteness. We follow two different ways of specifying our regression equation for villages located close to district borders.

We begin by constructing grid cells each side of which is one-fifteenth of a degree of latitude or longitude. There are 48,037 such grid cells across India. We assign each village to a grid cell, and retain only those grid cells that cross a district border. We are left with 6,057 grid cells, and 70,061 villages that are located close to district borders.

Within a grid cell, distance to nearest town, distance to nearest highway and local population density change smoothly, but there is a discontinuous jump in distance to district headquarters across the district boundary. We use this variation to identify the effect of administrative remoteness on a range of public goods and economic outcomes.

We estimate the following equation:

$$\mathrm{y}_{v,d,c} = \beta_0 + \beta_1 DistHQ + \beta_2 DistTown + \beta_3 DistHighway + \zeta Density_{v,j} + \mu_c + \eta_d + \epsilon_{v,d,c} \quad (1)$$

where $y_{v,d,c}$ is the outcome of interest for village v in district d and gridcell c. $DistHQ$ is the geodesic distance in kilometers from village v to its district headquarters. $DistTown$ is geodesic distance in kilometers from village v to the nearest town with population greater than 10,0000. $DistHighway$ is distance of village v to the nearest highway. $Density_{v,j}$ is the local population density in persons per square kilometer within a j kilometer radius of the village. In our regressions, we control for densities up to a distance of 15 kilometers. μ_c is the grid cell fixed effect, η_d is district fixed effect.

4 Results

In this section, we describe and discuss the main results (Section 4.1), robustness (Section 4.2) and the evidence on the mechanism (Section 4.3). We first show that distance from district headquarters 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 size of our comparison area, or the manner in which we the geographical location of the villages enters the regression specification. We then consider an important mechanisms that could explain these results, finding that at this stage, the evidence best supports higher cost of constructing public assets in these villages.

4.1 Main results

We begin by estimating correlations between rural outcomes (public goods provision as well as economic outcomes) and administrative remoteness for the full sample of our villages using an Ordinary Least Squares (OLS) regression. In each regression, we control for local population density and use district fixed effects. Table 2 reports these estimates for a range of public goods as reported in the 2011 Population Census. We note, for example, that within a district, distance from district headquarters - our measure of administrative remoteness - is negatively correlated with the probability of receiving a paved road, electricity, primary and secondary school, and health center.

Table 3 reports correlation estimates between administrattive remoteness and rural economic outcomes from both the 2011 Population Census as well as the 2012 SECC. We note that within a district, distance from district headquarters is negatively correlated with average rural income, share of households in the village with a solid roof, share of literates in the village population, share of village workforce engaged in nonfarm activities and share of agricultural land that is irrigated. It is positively correlated with the share of households that reported a migrant household member in the 2002 BPL Census.

As we explain in Section 3, we do not expect the OLS regression to identify the causal effects of administrative remoteness on rural outcomes. Hence we retain only those villages that are ocated close to district borders. We do so by constructing equally-spaced grid cells with each side equal to one-fifteenth of a degree, assigning villages to grid cells, and retaining only those grid cells that cross district boundaries.

Table 4 presents estimates from Equation 3 for the effect of distance from district headquarters on rural public goods provision. We note that while the probability of receiving electricity and having a primary school or a health center does not vary systematically within the grid cell, villages face considerable cost of administrative remoteness in terms of reduced access to paved roads and to secondary school. Since the standard deviation of distance to district headquarters 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 1.4% and that the village has a secondary school by 6.2%.

Table 5 presents estimates from Equation 3 for the effect of distance from district headquarters on rural economic outcomes. Several regression coefficients are statistically significantly different from zero using 99% confidence intervals. A one standard deviation increase in distance from district headquarters reduces average monthly income by 1.1% and the share of households in the village reporting a solid roof by 1.4%. At the same time, the proportion of village residents who are literate decreases by 0.8% and the proportion of workforce engaged in nonfarm activities by 1.4%. Crucially, distance to district headquarters does not matter for the share of households reporting at least one migrant family member in 2002, but distance to nearest town with population greater than 10,000 does matter. A one standard deviation increase in distance to nearest town increases the share of households with atleast one migrant family member by 4% on average.

Figure 3 shows the coefficient plot comparing 2011-2012 regression estimates (from Equation 3) with estimates from 1991. Over the 20 year period, the cost of administrative remoteness for literacy rate, nonfarm employment, and irrigation has remained fairly consistent (though in 2011, we cannot reject that the coefficient for irrigation is equal to 0 using a 95% confidence interval).

The administrative remoteness penalty for paved road access has reduced considerably over the 20 year period. This may be due to the large rural roads program (PMGSY) that the federal government started implementing in the early 2000s (Asher and Novosad, 2015). For secondary schools, the reverse seems to have happened. While we cannot reject the hypothesis that the 1991 coefficient for secondary school is equal to zero, there is a significant penalty in 2011. This may be due to a concentration of new secondary school construction in villages closer to district headquarters.

These diverse changes over the 20 year period can be seen most clearly in Table 6, which reports regression estimates from Equation 3, with the outcome variables expressed in terms of changes between 1991 and 2011 rather than in 2011 (or 1991) levels. The most striking estimate is for the change in probability of paved road access during 1991-2011. A one standard deviation increase in distance to district headquarters increases the probability that an unconnected village receives a paved road by almost 3.5%. This points towards the benefits of the PMGSY program being biased positively towards more administratively remote locations.

4.2 Robustness

In this section, we show the robustness of our results to changing the size of the grid cells, from one-fifteenth of a degree (approximately 7.4 kilometers) to double that size (approximately 14.8 kilometers per side). Tables 7 and 8 show the results of estimating Equation 3 for larger grid cells for public goods provision and economic outcomes respectively. We find that qualitatively our results remain the same, though a larger sample size helps us detect the effects of administrative remoteness with more precision.

4.3 Mechanism

The results we observe can be explained by several mechanisms. Villages located at farther distances from the district headquarters 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 headquarters, 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 preliminary 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 (Asher and Novosad, 2015) to say something about how unit costs change as a function of distance to district headquarters.

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 9 reports estimates from regressing PMGSY variables on distance to district headquarters for villages located close to district borders. We find that a one standard deviation increase in distance to district headquarters increases the cost per kilometer by 1.54%. This provides evidence for the first channel through which administrative remoteness may affect rural public goods provision in

India.

5 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, demographic 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 headquarters 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 headquarters 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 headquarters. 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 headquarters. 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	Closer Villages	Remoter Villages
Distance to District HQ (kms)	38.28	20.18	56.51
	(23.92)	(8.049)	(20.49)
Distance to nearest town (kms)	15.83	13.06	18.63
	(10.67)	(7.072)	(12.75)
Population (2011)	1484.6	1579.0	1406.5
	(2018.7)	(2182.6)	(1849.5)
Mean monthly earnings (2012 Rupees)	5113.7	5203.0	4758.6
	(2451.6)	(2173.7)	(1961.9)
Percent households with solid roof (2012)	47.75	52.30	42.73
	(34.85)	(33.91)	(34.96)
Percent population literate (2011)	57.27	58.61	55.75
	(13.94)	(13.59)	(14.06)
Percent workforce in nonfarm activities (2011)	27.86	32.25	23.26
	(26.78)	(27.96)	(24.53)
Percent villages electrified (2011)	61.87	60.36	63.17
	(48.57)	(48.91)	(48.23)
Percent villages with govt primary school (2011)	83.89	82.37	85.91
	(36.76)	(38.10)	(34.80)
Percent villages with health center (2011)	22.90	20.49	25.52
	(42.02)	(40.36)	(43.60)
Percent land irrigated (2011)	57.71	61.57	53.93
	(38.34)	(38.32)	(37.91)
Paved Road Access (2011)	80.50	81.54	79.51
	(39.62)	(38.80)	(40.36)
Distance to nearest highway (kms)	8.944	7.079	10.84
	(8.186)	(5.920)	(9.590)
Observations	395184	195811	195218

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 headquarters is less than distance to district headquarters for the median village. The "remoter villages" column presents values for villages whose distance to their district headquarters is larger than distance to district headquarters for the median village.

Table 2: OLS for public goods provision

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.027	-0.036	-0.005	-0.020	-0.016
	(0.003)***	(0.003)***	(0.003)*	(0.003)***	(0.003)***
Distance to nearest town (kms)	-0.121	-0.201	-0.013	-0.060	-0.028
	(0.007)***	(0.007)***	(0.007)*	(0.007)***	(0.008)***
Distance to nearest highway (kms)	-0.133	-0.183	-0.063	-0.087	-0.094
	(0.008)***	(0.008)***	(0.008)***	(0.008)***	(0.009)***
Outcome Mean	79.99	60.4	83.53	16.05	22.68
Fixed effects	District	District	District	District	District
Density controls	Yes	Yes	Yes	Yes	Yes
N	405706	405712	405699	405712	405712
R2	.3355	.4752	.176	.1035	.2022

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

Notes: The table presents regression estimates from Equation 3, where we regress public goods provision on distance to district headquarters in kilometers. However, we show these results for the full sample of villages (not just villages near district borders) and we do not include grid cell fixed effects. Robust standard errors are reported below point estimates.

Table 3: OLS for economic outcomes

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Households with a migrant
Distance to District HQ (kms)	-1.629	-0.076	-0.034	-0.048	-0.074	0.037
	(0.151)***	(0.002)***	(0.001)***	(0.002)***	(0.002)***	(0.003)***
Distance to nearest town (kms)	-8.963	-0.134	-0.101	-0.133	-0.190	0.116
	(0.365)***	(0.004)***	(0.002)***	(0.005)***	(0.005)***	(0.007)***
Distance to nearest highway (kms)	-8.498	-0.200	-0.127	-0.188	-0.006	0.087
	(0.421)***	(0.005)***	(0.002)***	(0.005)***	(0.006)	(0.008)***
Outcome Mean	4956	47.04	56.96	27.94	57.66	57.76
Fixed Effects	District	District	District	District	District	District
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	405712	405680	405174	402826	393591	283695
R2	.2751	.6464	.4738	.3194	.6225	.3883

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

Notes: The table presents regression estimates from Equation 3, where we regress economic outcomes on distance to district headquarters in kilometers. However, we show these results for the full sample of villages (not just villages near district borders) and we do not include grid cell fixed effects. Robust standard errors are reported below point estimates.

Table 4: Administrative remoteness and public goods, using small grid cells

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.046	-0.015	0.020	-0.038	-0.010
	(0.015)***	(0.014)	(0.016)	(0.015)**	(0.016)
Distance to nearest town (kms)	-0.095	-0.060	-0.159	0.014	-0.083
	(0.075)	(0.072)	(0.079)**	(0.077)	(0.082)
Distance to nearest highway (kms)	-0.152	-0.189	-0.113	-0.166	-0.241
	(0.079)*	(0.076)**	(0.083)	(0.081)**	(0.087)***
Outcome Mean	79.22	58.9	82.71	14.58	20.42
Fixed effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density controls	Yes	Yes	Yes	Yes	Yes
N	64245	64246	64245	64246	64246
R2	.4353	.6439	.2794	.215	.3063

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

Notes: The table presents regression estimates from Equation 3, where we regress rural access to public goods on distance to district headquarters in kilometers using grid cells with each side equal to one-fifteenth of a degree of latitude and longitude. All outcome variables reported here are binary variables that take the value 1 if the village has the public good in the 2011 Population Census, and 0 otherwise. Robust standard errors are reported below point estimates.

Table 5: Administrative remoteness and economic outcomes, using small grid cells

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Households with a migrant
Distance to District HQ (kms)	-2.178	-0.027	-0.018	-0.043	-0.016	-0.001
	(0.761)***	(0.008)***	(0.004)***	(0.009)***	(0.009)*	(0.015)
Distance to nearest town (kms)	-0.266	0.020	-0.073	-0.164	-0.128	0.250
	(3.847)	(0.040)	(0.021)***	(0.047)***	(0.045)***	(0.072)***
Distance to nearest highway (kms)	-19.912	-0.262	-0.142	-0.448	-0.113	0.090
	(4.068)***	(0.042)***	(0.022)***	(0.049)***	(0.048)**	(0.076)
Outcome Mean	4943	46.47	56.59	25.92	58.66	59.57
Fixed Effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	64246	64241	64171	63683	62425	43889
R2	.4516	.7876	.6693	.4796	.7752	.5579

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

Notes: The table presents regression estimates from Equation 3, where we regress economic outcomes on distance to district headquarters in kilometers using grid cells with each side equal to one-fifteenth of a degree of 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. Robust standard errors are reported below point estimates.

Table 6: Change in public goods 1991-2011

	Paved road	Electrification	Primary school	Secondary school	Medical center
Distance to District HQ (kms)	0.064	0.010	0.036	-0.025	-0.049
	(0.022)***	(0.019)	(0.019)*	(0.014)*	(0.019)**
Distance to nearest town (kms)	0.261	0.002	0.077	0.065	0.001
	(0.110)**	(0.099)	(0.095)	(0.070)	(0.096)
Distance to nearest highway (kms)	0.779	0.063	-0.121	-0.072	-0.271
	(0.117)***	(0.104)	(0.100)	(0.074)	(0.101)***
Outcome Mean	43.75	32.5	10.97	7.526	-13.15
Fixed Effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density Controls	Yes	Yes	Yes	Yes	Yes
N	64245	64246	64245	64246	64246
R2	.3394	.468	.1873	.1765	.5125

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

Notes: The table presents regression estimates from Equation 3, where we regress the change in public goods provision on distance to district headquarters in kilometers using grid cells with each side equal to one-fifteenth of a degree of latitude and longitude.. Each outcome variable is the change in probability that the village has the corresponding public good between 2011 and 1991. For example, "paved road" is the change in probability that the village has a paved road between the 1991 Population Census and the 2011 Population Census. Robust standard errors are reported below point estimates.

Table 7: Administrative remoteness and public goods, using large grid cells

	Paved Roads	Electrification	Primary School	Secondary School	Medical Center
Distance to District HQ (kms)	-0.033	-0.004	0.005	-0.028	0.013
	(0.010)***	(0.010)	(0.010)	(0.010)***	(0.011)
Distance to nearest town (kms)	-0.106	-0.066	-0.050	0.019	-0.058
	(0.028)***	(0.028)**	(0.029)*	(0.029)	(0.031)*
Distance to nearest highway (kms)	-0.254	-0.169	-0.072	-0.144	-0.263
	(0.030)***	(0.030)***	(0.032)**	(0.031)***	(0.034)***
Outcome Mean	79.12	59.19	82.81	15.2	21.16
Fixed effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density controls	Yes	Yes	Yes	Yes	Yes
N	147505	147507	147505	147507	147507
R2	.3845	.5781	.2291	.1505	.2484

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

Notes: The table presents regression estimates from Equation 3, where we regress rural access to public goods on distance to district headquarters in kilometers using grid cells with each side equal to 1/7.5 of a degree of latitude and longitude. All outcome variables reported here are binary variables that take the value 1 if the village has the public good in the 2011 Population Census, and 0 otherwise. Robust standard errors are reported below point estimates.

Table 8: Administrative remoteness and economic outcomes, using large grid cells

	Mean Income	Solid Roof	Percent Literate	Percent Nonfarm	Percent Land Irrigated	Households with a migrant
Distance to District HQ (kms)	-1.875	-0.023	-0.024	-0.040	-0.022	0.017
	(0.499)***	(0.005)***	(0.003)***	(0.006)***	(0.006)***	(0.009)*
Distance to nearest town (kms)	-2.802	-0.121	-0.047	-0.169	-0.096	0.105
	(1.456)*	(0.015)***	(0.008)***	(0.018)***	(0.018)***	(0.027)***
Distance to nearest highway (kms)	-11.277	-0.185	-0.152	-0.307	-0.069	0.077
	(1.574)***	(0.017)***	(0.009)***	(0.019)***	(0.019)***	(0.029)***
Outcome Mean	4942	46.85	56.63	26.25	59.03	59.1
Fixed Effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	147507	147492	147313	146378	143515	101115
R2	.375	.7458	.6035	.4094	.7347	.4854

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

Notes: The table presents regression estimates from Equation 3, where we regress economic outcomes on distance to district headquarters in kilometers using grid cells with each side equal to 1/7.5 of a degree of 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. Robust standard errors are reported below point estimates.

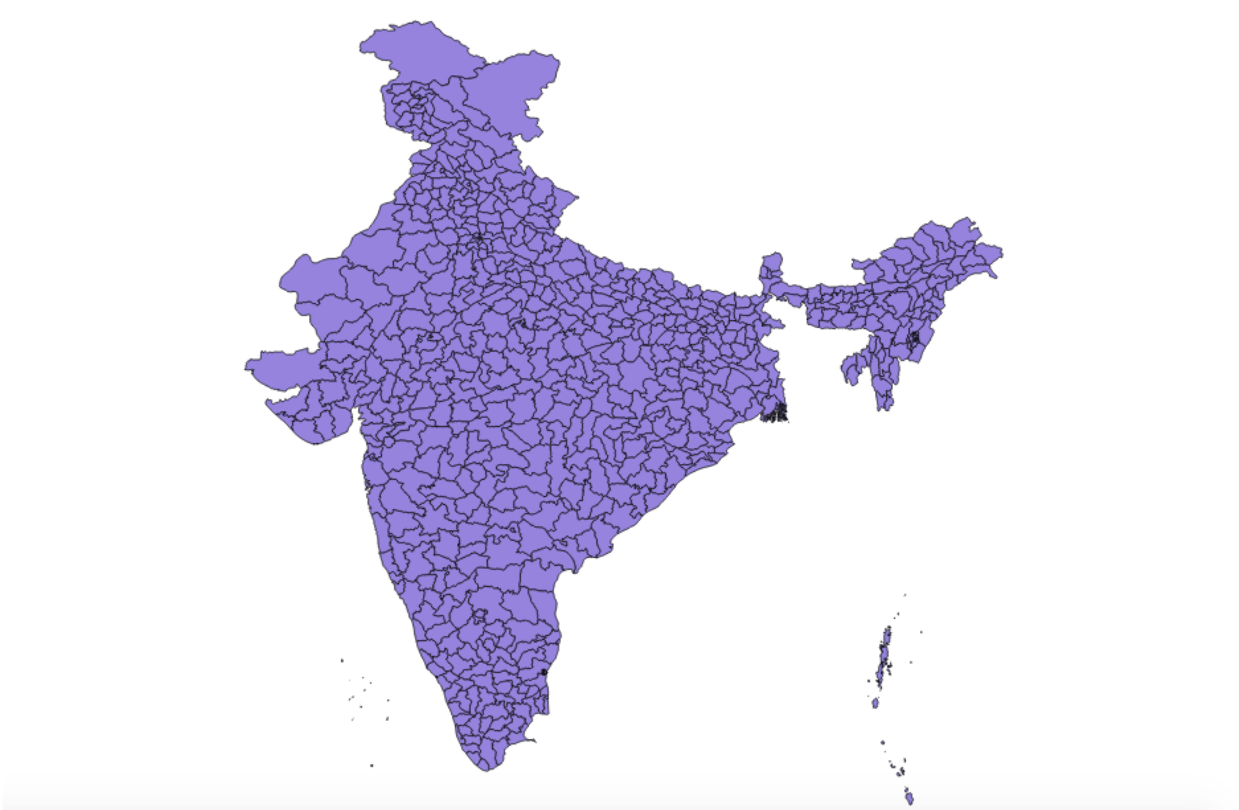
Table 9: PMGSY road construction costs

	Cost Per km	Cost Overrun Per km	Time Overrun Per km	Time Per km
Distance to District HQ (kms)	0.003	-0.001	0.245	0.439
	(0.002)*	(0.001)	(0.287)	(0.365)
Distance to nearest town (kms)	0.006	-0.007	-0.110	-1.256
	(0.007)	(0.004)	(1.319)	(1.668)
Distance to nearest highway (kms)	-0.013	-0.005	1.012	0.338
	(0.007)*	(0.004)	(1.357)	(1.714)
Outcome Mean	3.166	-.1916	83.69	236.2
Fixed effects	Grid-cell, District	Grid-cell, District	Grid-cell, District	Grid-cell, District
Density controls	Yes	Yes	Yes	Yes
N	11384	8550	9578	9565
R2	.7771	.5917	.5614	.5614

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

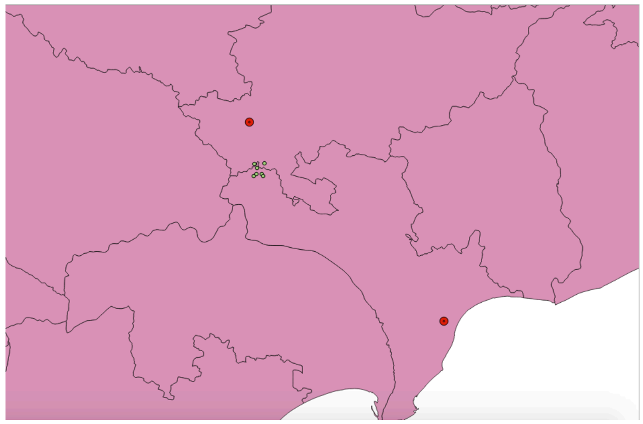
Notes: The table presents regression estimates from Equation 3, where we regress PMGSY project variables on distance to district headquarters in kilometers using grid cells with each side equal to 1/7.5 of a degree of latitude and longitude. Cost per kilometer is the final cost of constructing the PMGSY 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. Robust standard errors are reported below point estimates.

Figure 1: Map of Indian Districts



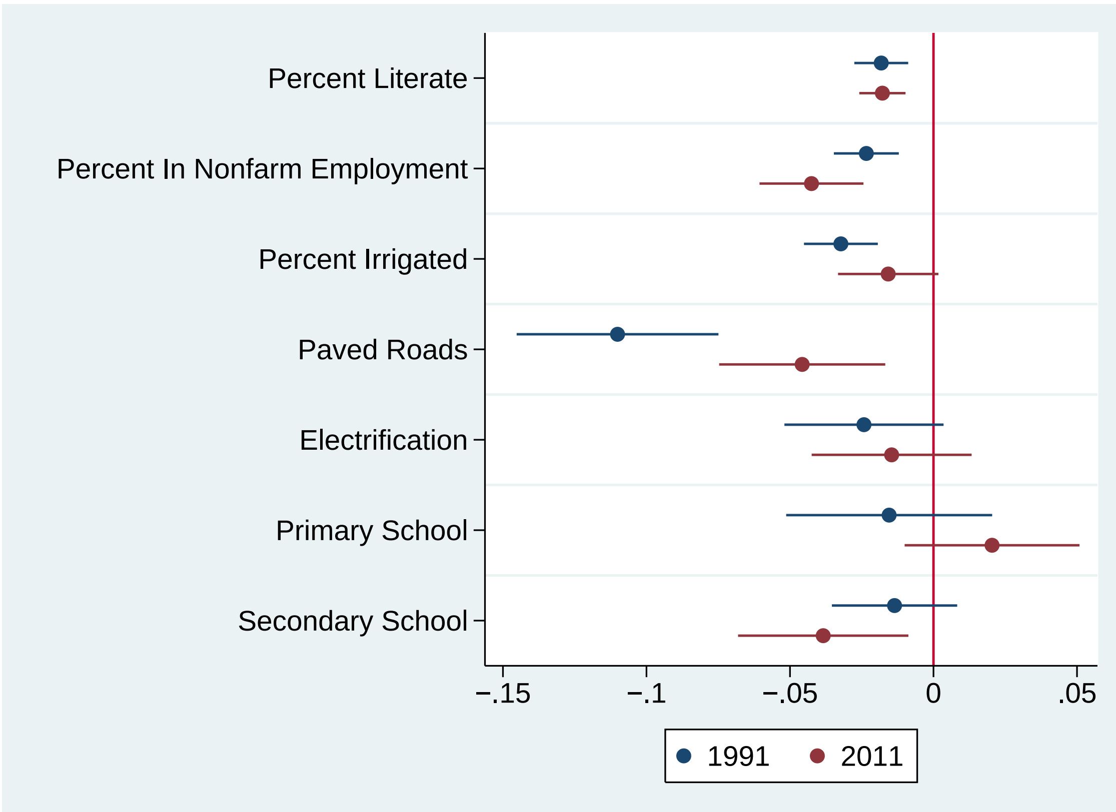
Notes: The map shows all Indian districts during the period of our study.

Figure 2: Khammam-Krishna District Border



Notes: Example of our RD strategy. The villages lying on the northern side of the district border, in Khammam district, are on average about 20 kilometers from their district headquarters (the top red dot). Villages lying in the southern side of the district border, in Krishna district, are about 120 kilometers from their district headquarters (the bottom red dot).

Figure 3: Comparing effects of administrative remoteness over time



Notes: Comparing regression coefficients in 1991 and 2011.

References

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

Asher, Sam and Paul Novosad, "The Employment Effects of Road Construction in Rural India," 2015.

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

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.

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

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.

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.

World Bank, "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.