The Cost of Distance: Geography and Governance in Rural India*

Latest version of the paper is available here

Sam Asher† Karan Nagpal‡ Paul Novosad§

February 23, 2018

Abstract

Increasing the effectiveness of the state is a major challenge facing most developing countries today. In this paper, we focus on one important factor that constrains the state's ability to provide public goods to all citizens: citizens' physical remoteness from their administrators. Using rich data on village India, and a spatial regression discontinuity design, we show that greater distance to administration reduces a village's access to public goods and worsens economic outcomes. Villages that are more remote from their administrators have fewer roads, schools, health centers and less irrigation. In turn, their residents have lower incomes, fewer assets, less literacy and are more likely to be employed in agriculture. Households in these villages are less likely to own enterprises. At least for roads, these effects are not driven by the higher cost of construction in remote villages, but higher cost of monitoring road quality. Our results suggest that reducing the distance between the state and its citizens and changing bureaucrat incentives can help to mitigate the large spatial disparities in living standards observed within many developing countries.

JEL Codes: R12, D63, H41, O18

*This version November 2017. We thank seminar participants at the Center for Global Development (Washington DC), the DFID-IZA Workshop (Oxford), the Indian Statistical Institute Annual Conference (Delhi), the CSAE Conference (Oxford), the Nottingham PhD Students Workshop, the IEA Congress (Mexico City), the Indian Ministry of Finance (Delhi), the National Council for Applied Economic Research (Delhi) and the National Institute for Public Finance and Policy (Delhi). We are grateful for discussions with James Fenske, Doug Gollin, Beata Javorcik and Simon Quinn. We are indebted to Taewan Roh and Kathryn Nicholson for exemplary research assistance.

†The World Bank

‡University of Oxford. Address: Jesus College, Turl Street, Oxford OX1 3DW, United Kingdom. Email: karan.nagpal@economics.ox.ac.uk

§Dartmouth College

1 Introduction

In many developing countries, the state has committed to provide public goods such as roads, schools, and health facilities to all citizens. Yet many of them still lack access to these amenities (The World Bank, 2015), even when sufficient resources are available to provide them (Mookherjee, 2015). What constrains the effectiveness of the state in providing universal access to public goods? The economics literature has examined the constraints on both the state's ability to supply public goods (Banerjee et al., 2013; Finan et al., 2015),[1] as well as citizens' ability to demand them (Bardhan, 1984; Alesina et al., 2003; Banerjee et al., 2007).[2]

The state's capacity may be further restricted by the geography of its administration, in particular, by the physical distance between administrators and citizens. The trade literature has documented the importance of transport costs and information frictions in restricting the flow of private economic activities over space (Adamopoulos, 2011; Gollin and Rogerson, 2014; Allen, 2014; Atkin and Donaldson, 2015). Do these frictions also restrict the flow of public services over space? This has been an important motivation for recommending decentralization as a way of improving public service delivery in developing countries (Bardhan, 2002; Mansuri and Rao, 2013). However, testing this hypothesis is empirically challenging: administrative headquarters, where public officials are based, are often located in towns that also have large goods and labor markets. Distance to administration is therefore correlated with distance to important markets, making it difficult to isolate its effect on rural outcomes.

In this paper, we test whether distance to administration, what we refer to as "administrative remoteness," matters for the provision of public goods and services, and consequently, whether it affects economic development. We do this by focussing on village distance to district headquarters in India. To recover causal estimates, we employ a spatial regression

[1] See, for example, Hanna and Wang (2017) on the type of individuals who enter the civil service; Muralidharan and Sundararaman (2011) and Ashraf et al. (2014) on the incentives they face; and Rasul and Rogger (forthcoming) on how the state manages its civil servants.

[2] On the pernicious effects of elite capture, patronage and clientelism, see, for example, Callen et al. (2013), Anderson et al. (2015), or Burgess et al. (2015).

discontinuity design that compares outcomes for villages located on either side of district borders. These villages are similar in all key respects, but they have substantially different distances to their administrative headquarters. We exploit this variation to examine how – and why – administrative remoteness affects access to public goods, and how these differences, in turn, shape the spatial distribution of poverty and prosperity in rural India.

To do this analysis, we construct a panel dataset covering 600,000 Indian villages, with detailed information on the availability of public goods in each village. We match this to data from a national census that provides household-level information on incomes, assets, and employment structure for 800 million rural individuals. Further, we obtain geocoordinates for all villages, towns and cities in India. We use them to construct measures of administrative remoteness for all Indian villages, as well as to calculate their distance to towns and cities. The high spatial resolution and comprehensive coverage enable us to investigate the generalized effects of administrative remoteness in a way that has not been possible before through more aggregate surveys.

We find that administrative remoteness worsens rural access to several public goods that are provided by district administrations in India. Villages that are more distant from their administrative headquarters are less likely to have paved roads, primary schools, health centers, and major irrigation facilities. In turn, a smaller share of their workforce is engaged in the non-farm sector.[3] A lower share of households in these villages report owning enterprises and relying on regular salaried employment. These differences in the employment structure have a bearing on rural living standards. For example, administratively remote villages produce a lower intensity of nightlights as compared to neighboring villages that are closer to their administrative headquarters. Average monthly income is also lower in these villages.[4]

[3]Rural employment structure affects rural productivity, because labor productivity in agriculture is lower than in non-farm sectors of the rural economy (Gollin et al., 2014; Restuccia et al., 2008; Caselli, 2005).

[4]We construct a measure of avergae village monthly income using data from the Socioeconomic and Caste Census 2012 and the National Sample Survey 2011-12, as explained in Section 3.2.

Further, their residents have worse quality housing, fewer household assets, and lower rates of literacy.

One of the channels driving these results is lower willingness or ability of public officials to visit more distant villages. While there are several ways in which public officials can obtain information about the village without travelling there, for some tasks, such as inspecting the quality of public infrastructure, there are few alternatives to physical visits. We confirm this hypothesis using data from India's national rural roads program. We first show that administratively remote villages are no less likely to receive roads under the program, and these roads do not cost more, or take longer to construct, than similar roads in neighboring villages that are closer to their headquarters. We then investigate whether the quality of the road was audited by national quality monitors employed under the program. This monitoring was intended to be randomly distributed across beneficiary villages. If distance to administration does not matter, we should find no systematic differences in monitoring between administratively remote and proximate villages. However, roads constructed in remote villages are 9.4 percent (2.2 percentage points) less likely to be audited under the program, confirming our hypothesis that high transport costs inhibit the state's ability to monitor infrastructure quality in more distant villages.

We consider and rule out four alternative hypotheses. First, governance quality may be correlated with district features such as geographical area, nature of colonial institutions,[5] or the level of federal budget support.[6] To account for any factor that changes at the level of the district, we employ district fixed effects throughout our analysis. Second, differential outcomes in remote villages may be driven by other correlates of remoteness that reduce overall access to public goods. We show that administratively remote villages are no less

[5]Colonial institutions such as land revenue systems have been shown to affect contemporary public goods provision (Banerjee and Iyer, 2005; Banerjee et al., 2005; Iyer, 2010)

[6]The erstwhile Planning Commission had classified about 250 districts as "backward", which made them eligible for additional support in federal programs.

likely to receive public goods provided by other tiers of administration, such as the federal government or village councils. Third, the economic effects of administrative remoteness could be explained by sorting across the district borders, as residents of administratively remote villages relocate to the "better" side of the border. We do not find any evidence of such relocation, consistent with the literature on rural migration in India.[7] Finally, provincial borders may restrict rural citizens' access to economic opportunities.[8] We show that our results are robust to dropping state borders from the sample.

These results have implications for several strands of the literature. First, these results contribute a new dimension to our understanding of spatial inequality in developing countries. One of the questions investigated by this literature is why remoteness from urban centers is associated with worse economic outcomes. Several papers have documented the negative effects of greater distance to towns, which host large goods and labor markets, on poverty, gains from trade, and structural transformation (Fafchamps and Wahba, 2006; Feyrer, 2009; Michaels et al., 2012; Atkin and Donaldson, 2015; Storeygard, 2016). In this paper, we argue that towns do not just provide large markets for goods and labor, but also serve as administrative centers. We show that greater distance to such towns reduces rural access to public services, providing another channel through which remoteness from towns worsens economic outcomes.

Further, one of the central premises for decentralization is that reducing the distance between the state and its citizens can help to improve the efficiency of fiscal expenditure (Oates, 1972; Bardhan, 2002). It has been difficult to test this premise empirically. While there is a growing empirical literature on the benefits of decentralization (Faguet, 2004; Barankay and Lockwood, 2007; Galiani et al., 2008; Kis-Katos and Sjahrir, 2017), these episodes often

[7] See, for example, Munshi and Rosenzweig (2016).

[8] For example, Kone et al. (2017) show that the lack of portability of welfare benefits, as well as provincial-level biases in admission to tertiary educational institutions and public sector recruitment, restrict migration across provincial borders in India.

accompany widespread political change in the country, making it difficult to disentangle the different components of decentralization. Our empirical framework allows us to isolate the pure effects of lower distance between citizens and the state, contributing causal evidence to support one of the central premises of decentralization.

Relatedly, there is a growing literature on the politics of administrative unit proliferation (Grossman and Lewis, 2014; Pierskalla, 2016; Grossman et al., 2017). This literature attributes the surge in the number of sub-national administrative units in developing countries to political economy factors such as creating administrative jobs for political supporters (Green, 2010; Hassan, 2016), ethnic patronage, and clientelism. Our results show that apart from their political consequences, smaller administrative units can also help improve the delivery of public services and therefore enhance state capacity in developing countries.

Finally, we contribute to a long-standing literature on the provision of public goods in developing countries. This literature has adopted two broad approaches to explain how citizens get access to public goods: bottom-up processes, in which communities organize to demand public goods from the state (Banerjee and Iyer, 2005; Dell, 2010; Anderson et al., 2015; Dell et al., 2017), and top-down interventions, in which the state expands access to public goods throughout the country (Duflo, 2001; Banerjee et al., 2007; Iyer, 2010; Burgess et al., 2015). In this paper, we provide evidence for a significant channel through which both top-down interventions work: distance to administration. Greater distance increases costs to the state of supplying the public goods and monitoring their quality. It may also increase costs for communities to organize and demand public goods from the administration.

In the next section (Section 2), we explain the context of public administration in India. Then in Section 3, we describe the multiple data sources and explain the construction of our main variables. Section 4 outlines our empirical strategy. Section 5 presents our main results. Section 6 discusses the mechanisms, and Section 7 concludes.

2 District administration in India

Public administration in India is divided into several tiers. At the top is the federal government, which is responsible for national public goods such as national security and diplomacy. The federal government also manages organizations that provide public goods through national networks, such as railways, electricity, postal services, and highways.[9]

Then there are state governments in India's twenty-eight states. State governments are responsible for managing public health and public education systems, along with a wide range of social and anti-poverty programs (Second Administrative Reforms Commission, 2009). For many social sectors, the federal and state governments work together: the federal government designs the schemes and provides a significant proportion of the funding, while the state governments take charge of implementing the schemes within their territories. For example, the national rural roads program, the Pradhan Mantri Gram Sadak Yojna (PMGSY), is a central government program that was launched in 2000. Through this program, the central government has provided state governments more than US$ 1.5 billion to construct approximately 400,000 kilometers of all-weather (or paved) access roads to connect more than 185,000 rural habitations to the state and national road network (Asher and Novosad, 2017).

States are further organized into districts. In 2012, India had 640 districts. The district administration is responsible for implementing all federal and state government projects in the district. Unlike the federal and state governments, where an elected executive commands authority over the bureaucracy, district administrations are almost entirely bureaucratic in character (Second Administrative Reforms Commission, 2009).[10] The district administration is headed by the District Collector,[11] who often belongs to the elite national cadre of civil servants, the Indian Administrative Service (IAS). The District Collector is considered

[9] The railways are managed by the Ministry of Railways, electricity through the Rural Electrification Corporation, postal services through India Post, and highways through the National Highway Authority of India.

[10] The 73rd and 74th Constitution Amendment Act created district councils, or *zila parishads*, but these have limited powers in most states, with Gujarat and Maharashtra being notable exceptions.

[11] The Collector is also known as the District Magistrate or the Deputy Commissioner.

by many as the chief representative of the Indian state within the district (Kothari, 1970) and has considerable discretion in allocation of central and state funds to different villages within the district. For example, in the PMGSY program, even though the central government specified fairly objective elegibility criteria for allocating paved roads to villages, it was the district administration that was responsible for designing the district PMGSY network, after consulting with elected representatives and taking into account special needs for marginalized communities (Asher and Novosad, 2017). To facilitate administration and monitoring at local levels, districts are further divided into sub-districts[12] and blocks.

In 1993, the Indian Constitution was amended to create village councils, called *Gram Panchayats*, to serve groups of villages all over the country. These are elected councils that receive central and state government funds to provide local public goods, such as footpaths and gravel (unpaved) roads, as well as water and sanitation facilities. The village councils are also responsible for identifying welfare beneficiaries for government programs, and for implementing the National Rural Employment Guarantee Act (Chattopadhyay and Duflo, 2004; Anderson et al., 2015).

3 Data

To examine the effects of administrative remoteness, we bring together several sources of administrative data identified at the village level. We also obtain geocoordinates for all towns and villages in India and use these to calculate our distance measures. Below, we describe each dataset in greater detail.

3.1 Population Census

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 village-level data from the Population Census 2011, which reports demographic characteristics such as popula-

[12] These are called *tehsils*, *talukas* and *mandals* in different states.

tion size, age and gender composition, literacy rate, and the share of village population that belongs to the Scheduled Castes (SC) and Scheduled Tribes (ST), which are traditionally marginalized communities in rural India. In addition, the Population Census provides rich information on the availability of various amenities such as different types of roads, schools, health centers, water and sanitation facilities, electricity, post offices, polling booths, and irrigation facilities in the village.

3.2 Socioeconomic and Caste Census

In 2012, the Government of India conducted the Socioeconomic and Caste Census (SECC), a one-off census of all households in the country, to collect detailed information on assets, incomes, occupation structure and demographic characteristics at the household level. This information is substantially richer than information collected during the Population Census. The Government of India made the SECC publicly available on the Internet in PDF and Excel formats. To construct a useful dataset from these files, we scraped the Excel and PDF files from the Government's SECC website, parsed the embedded text, and translated the data to English from twelve Indian languages (Asher and Novosad, 2017).

In this paper, we use household level data from the SECC. This includes monthly income of the highest earning member of the household (divided into 3 categories: less than INR 5,000 per month, between INR 5,000 and 10,000 per month, and more than INR 10,000 per month), important sources of earnings, measures of housing quality, and the household's ownership of different types of assets. We use the household-level data to generate village averages, which we match to village data from the Population Census.

To generate a proxy for average village income, we combine the SECC income data with consumption expenditure data from the 68th round (2011-12) of the National Sample Survey (NSS). We use the NSS data to generate conditional averages of household income. For example, conditional on monthly income being less than INR 5000 per month, the aver-

age household income is INR 3,076. Conditional on monthly income being between INR 5,000 and 10,000 per month, the average household income is INR 6,373, and conditional on monthly income being greater than INR 10,000, the average household income is INR 22,357. We use these conditional averages from the NSS to calculate a proxy for average village income. This technique provides us a measure of average monthly income for every village in India, which is a unique contribution from this dataset.

3.3 GIS data

We obtained village and town geocoordinates, GIS data on current and historic district borders, and GIS data on rivers from ML Infomap. We obtained highway GIS data from both OpenStreetMap and the National Highways Authority of India.

4 Empirical Strategy

Isolating the effects of administrative remoteness is challenging because administrative headquarters are often located in the largest towns in the district. Therefore, conventional measures of market access – such as distance to towns, cities and highways – are correlated with distance to administrative headquarters. Therefore, our empirical strategy focusses on villages located close to district borders. We show that while distance to towns and cities, and demographic and geographic variables change smoothly across the border, distance to administrative headquarters changes sharply at the border.

We illustrate the empirical strategy in Figure 1. The polygons in the figure are two districts in the western state of Maharashtra: Ahmadnagar and Beed. Their district headquarters are represented by black diamonds. We focus on the villages close to the Ahmadnagar-Beed border. This border is quite long, approximately 180 kilometers. To ensure that our comparisons are within neighboring villages, we create equally-spaced segments along the border. In this case, each segment is 20 kilometers long. Each color represents a different border segment. Within each segment, a filled circle represents a village located on the side of the

border that is closer to its administrative headquarters, while a hollow circle represents a village located on the side of the border that is more distant to its headquarters.

As we can observe, the northern and southern ends of the border are closer to the district headquarters of Beed. Therefore, the villages on the Beed side of the border are represented by filled circles, and villages on the Ahmadnagar side are represented by hollow circles. However, the middle section of the border is closer to the district headquarters of Ahmadnagar. Therefore, the villages on the Ahmadnagar side are now represented by filled circles, and villages on the Beed side are represented by hollow circles. Our identifying variation comes from a change in administrative remoteness along the same district border, i.e. within the same district-district dyad.

Following Dell (2010), we estimate the following spatial regression discontinuity equation:

$$\begin{aligned} \mathrm{y}_{v,d,s} = \beta_1 RemoteSide_{d,s} + \beta_2 DistTown_v + f(\mathrm{Location}_v) \\ + \beta_3 Geograph_v + \beta_4 Demograph_v + \delta_d + \eta_s + \epsilon_{v,d,s} \end{aligned} \tag{1}$$

where $y_{v,d,s}$ is the outcome of interest for village v located in district d, along segment s of the border between district d and the adjoining district. These outcomes include access to different types of public goods and rural economic indicators, as described in Section 3.

RemoteSide is our binary treatment variable. For each segment s, we have villages on both sides of the district border. For each segment-district, we find the average distance to the corresponding district headquarters. The side that has the higher average distance to headquarters is "administratively remote", and hence $RemoteSide_{d,s}$ takes the value 1 for all villages in this segment-district group. Villages on the other side are closer to their administrative headquarters, and hence for them, $RemoteSide_{d,s}$ is 0.

DistTown represents a set of controls for village's (geodesic) distance in kilometers to the nearest town above different population thresholds. We control for distance to towns whose

population in 2011 exceeded 10,000, 50,000, 100,000, and 500,000.

$f(\text{Location}_v)$ is the RD polynomial, which controls for smooth functions of geographic location. Following Gelman and Imbens (2017) and Dell and Olken (2017), $f(\text{Location}_v)$ is a local linear polynomial in latitude and longitude, estimated separately for each administrative border. These linear controls account for any spurious linear trends in outcomes at the border. The results are robust to using other types of RD polynomials, such as a polynomial that includes an interaction term between latitude and longitude, a quadratic polynomial in latitude and longitude, as well as polynomials that include village's distance to district border.

Geograph refers to geographic controls, such as mean elevation in the village and the village's distance to a major river. *Demograph* refers to demographic controls, which include the percentage of village population in 2011 that belonged to Scheduled Castes (SC) or Scheduled Tribes (ST), which are traditionally marginalized communities in rural India.

δ_d is the district fixed effect. This controls for any district-specific factors that may affect rural economic outcomes or the provision of public goods. This could include precolonial and colonial institutions such as land revenue systems, or district-specific factors such as average governance quality and district area. The inclusion of the district fixed effect ensures that the coefficient of interest, β_1, captures the effect of distance from district headquarters that is over and above any district-wide average governance measure. Therefore, our results are not driven by more "administratively remote" villages being located in worse-administered districts.

η_s is the border segment fixed effect. District borders can be quite long, ranging from a few kilometers to more than 100 kilometers. We want to ensure that we are comparing villages located in close geographic proximity. Further, as in Figure 1, some portions of the border may be close to one headquarters, while other portions are closer to the other headquarters. Therefore, we create equally-spaced segments along the district border. Segment s is the border segment closest to village v.

$\epsilon_{v,d,s}$ is the error term. We cluster the standard errors at the district-segment level, since this is the level of treatment in our analysis (Abadie et al., 2017). The results are robust to allowing spatial correlation in the standard errors, following the method used in Burgess et al. (2016).

The validity of our regression discontinuity design requires a number of key assumptions. These assumptions in turn impose certain restrictions on our sample of villages. First, we compare villages located close to the district border. In our baseline specification, we use a bandwidth of 6 kilometers around the district border. Since there is no widely accepted optimal bandwidth for regression discontinuities in multi-dimensional space, we follow Dell and Olken (2017) and show robustness to several different bandwidths.

Second, we want to ensure that villages on either side of the district border have similar access to the closest town or highway. If this is not the case, the outcomes will be driven by a combination of distance to administration and distance to markets, and it would not be accurate to attribute them to administrative remoteness alone. Geographical features such as rivers, mountains and ridges create barriers (Nunn and Puga, 2012). Hence, we drop borders across which the change in elevation is greater than the 90th percentile (the results are robust to replacing the 90th percentile threshold with other thresholds, such as 80 and 99). We further show robustness to dropping administrative borders that coincide with rivers.

Third, we exclude states that are classified as "special category states" by the federal government. These are Jammu and Kashmir, Uttarakhand, Himachal Pradesh, Assam, Arunachal Pradesh, Manipur, Nagaland, Tripura, Meghalaya, Mizoram and Sikkim. Administration of these states faces special challenges such as "hilly and difficult terrain, low population density or the presence of sizeable tribal population, strategic location along international borders, economic and infrastructural backwardness and non-viable nature of State finances" (The Hindu, 2016). District borders in these states often coincide with geographical barriers such as mountains and ridges, limiting the validity of our identification

assumptions. Therefore, while our results are robust to including these states in the sample, we want to be conservative and exclude them from our main sample.

Finally, as a robustness exercise, we estimate the intensity of treatment equation, which uses distance to district headquarters to measure administrative remoteness, rather than the binary treatment variable:

$$\begin{aligned} \mathrm{y}_{v,d,s} = \beta_1 DistDistrictHQ_{v,d,s} + \beta_2 DistTown_v + f(\mathrm{Location}_v) \\ + \beta_3 Geographic_v + \beta_4 Demographic_v + \delta_d + \eta_s + \epsilon_{v,d,s} \end{aligned} \tag{2}$$

where $DistDistrictHQ_{v,d,s}$ is the geodesic distance in kilometers between village v and its district headquarters. All other variable definitions remain the same as in Equation 1.

4.1 Balance checks

For the control villages to be appropriate counterfactuals for the treatment villages, we require that all controls other than administrative remoteness should change smoothly at the border. If c_1 and c_2 are potential balance outcomes for remote and proximate villages in our sample respectively, then our empirical design requires that $E[c_1|x, y]$ and $E[c_2|x, y]$ vary continuously at the district border (Dell et al., 2017).

To test this, we regress each control variable on the treatment dummy, $RemoteSide_{d,s}$, and all the other controls in Equation 1. Estimates from these regressions are presented in Table 1. The first row shows the first stage estimate, from regressing distance to district headquarters on all the right hand side variables in Equation 1. The average treatment size, i.e. the average increase in distance to district headquarters as we move from the administratively proximate side of the border to the administratively remote side, is approximately 15.6 kilometers. This is equal to 37% of the mean distance to district headquarters in our sample. Figure 2 shows the discontinuous change in distance to district headquarters at the

border (the RD cutoff in our design).

The next four rows present estimates from regressing distance to the nearest town, whose population in 2011 exceeded a certain threshold, on the treatment dummy, $RemoteSide_{d,s}$ and all the other controls in Equation 1. If our empirical strategy is valid, we should not be able to reject the null hypothesis that β_1 in these regressions is zero. We find that we cannot reject the null hypothesis that β_1 is equal to zero for village distance to towns with population exceeding 10,000 and 500,000 in 2011. Further, the magnitude of the regression coefficients is miniscule as compared to the sample mean for these variables: 0.2% for distance to towns with population exceeding 10,000, and 0.03% for distance to towns with population exceeding 500,000.

However, treatment villages are marginally more distant to towns with population in 2011 exceeding 50,000 and 100,000. This is not surprising, since for most villages in our sample, the nearest town in this population range serves as a district headquarter, even though it may not be that village's district headquarter.

For example, consider the villages in yellow in Figure 1, located approximately in the middle section of the border. While treatment (administratively remote) and control (administratively proximate) villages are located very close to each other, given a bandwidth of 6 kilometers, they are still, on average, about 6 kilometers apart. For all these villages, the nearest large town is Ahmadnagar, with a population of 350,905 in 2011. For the control yellow villages, located on the Ahmadnagar district side of the border, the average distance to Ahmadnagar town is 18.5 kilometers. This is also the villages' average distance to their district headquarters. For the treatment yellow villages, located on the Beed district side of the border, the average distance to Ahmadnagar town is 22.9 kilometers, about 4.5 kilometers more than the average distance for control villages. However, the difference in treatment (average distance to district headquarters) is much larger, 89.6 kilometers for treatment villages in Beed versus 18.5 kilometers for control villages in Ahmadnagar. Hence, on average,

treatment villages are marginally more distant to towns with population greater than 50,000 or 100,000, given our empirical design. However, the magnitudes are a few hundred meters. For towns with population exceeding 50,000, the magnitude is 1.5% of sample mean, and for towns with population exceeding 100,000, the magnitude is 1.9% of sample mean.

To confirm that these statistically significant differences are caused by the average distance between treatment and control villages, we plot these coefficients for several different bandwidths in Figure 3 and Figure 4 (Figure 5 and Figure 6 do the same for the coefficient on distance to towns with population exceeding 10,000 and 500,000 respectively). These coefficient plots confirm that the average difference between treatment and control villages in distance to towns with population greater than 50,000 and 100,000 increases with the selected bandwidth around the border.

Figure 7 shows how distances to towns with different population sizes change across the district border. The graphs show that distances to towns are continuous at the border, except for the small discontinuous change in distance to towns with population greater than 50,000 and 100,000.

The next five rows of Table 1 show that demographic and geographic characteristics are also balanced for treatment and control villages, except for distance to nearest river. Treatment villages are 155 meters more distant to the nearest river, which is 0.3% of the mean distance to nearest river in our sample. We control for distance to nearest river in all our regressions, and also show robustness of the main results to dropping district borders that coincide with rivers. Figure 8 shows that non-distance controls are also continuous at the border, except for the small discontinuous change in distance to nearest river.

5 Results

Administrative remoteness reduces the provision of public goods, decreases the share of non-farm employment, and worsens poverty in rural India. In this section, we first present the

main effects on access to paved roads (Section 5.1), discussing robustness of the result to different specifications and sample restrictions. We then present results for access to different types of public goods, classified by the tier of administration that has the responsibility to provide them (Section 5.2), and discuss the economic effects of administrative remoteness (Section 5.3). Finally, we show robustness of these results in Section 5.4.

5.1 Roads

We first test whether administrative remoteness reduces the probability that a village has a paved road. Table 2 presents regression coefficients from estimating Equation 1 and Equation 2 for rural access to paved roads. We consider that a village has a paved road if the Village Directory of Population Census 2011 records that the village has an "all-weather road." Using our preferred specification, we find that villages on the administratively remote side of the border are 1.2 percentage points less likely to have a paved road than control villages. Given that 69% villages in our sample have a paved road, this implies a 1.7 percent reduction in paved road access relative to the sample mean. In terms of numbers of villages, 2,220 fewer villages have paved roads due to administrative remoteness, after accounting for distance to towns and cities and for district-level governance quality.

Since the binary treatment variable is an average of different treatment sizes, we also estimate Equation 2 in order to measure the impact of each additional kilometer in distance to the district headquarters. This result is presented in the last row of Table 2. Moving from the 25th to the 75th percentile of the distance to district headquarters distribution in our sample (from a distance of 25 kilometers to 53 kilometers) reduces the probability that a village has a paved road by 1.84 percentage points, or 2.6 percent of the sample mean.

Subsequent rows of Table 2 show robustness of the main result to different RD choices. First, Row 2 shows that the statistical significance changes marginally when we allow for spatial correlation in the standard errors. Then, in rows 3-5, we present estimates using

different RD polynomials: a linear polynomial in distance to border, a linear polynomial in distance to border and latitude and longitude, and a quadratic polynomial in latitude and longitude. The effect of administrative remoteness on paved road access is similar across all these regressions. Row 6 shows that the administrative remoteness effect is larger when we drop district borders that coincide with rivers, and when we drop state borders from our sample (Row 7). Finally, in Row 8, we observe that the main result is robust to restricting the bandwidth to 3 kilometers around the district border. Figure 9 presents the main result graphically, showing the drop in probability of paved road access for administratively remote villages, to the right of the boundary.

5.2 Other public amenities

Having established that administrative remoteness reduces rural access to paved roads, we now investigate whether it affects the provision of other public goods as well. Table 3 presents results for a wider set of public amenities, which we classify into 3 categories: those provided by the district administration, those provided by higher tiers of administration such as the federal government, and those provided by village councils.

First, the administrative remoteness effect is not limited to paved roads alone, but extends to other public goods for which the district administration is responsible. These include government primary schools, mobile health clinics, irrigation facilities, and treated tap water. Remote villages are about 1 percentage point less likely to have a government primary school, 0.3 percentage points less likely to have a mobile health clinic, and 1.1 percentage points less likely to have treated tap water. Further, the share of land irrigated in these villages is 0.7 percentage points lower. These results are robust to dropping state borders from our sample.

However, distance to district headquarters does not matter for access to public goods that are provided by the federal government, such as post offices, polling stations, national highways or electricity. Neither does it matter for access to local amenities for which the village

councils are responsible, such as unpaved (gravel) roads, and water and sanitation facilities. The fact that treatment and control villages are not different in their access to public goods provided by other tiers of administration suggests that treatment villages are not remote from all forms of administration. However, they lack paved roads, primary schools and other public goods due to administrative failure primarily on part of the district administration.

5.3 Economic impacts

We now test whether administrative remoteness, which reduces rural access to public goods, also affects economic outcomes. Table 4 presents results for four types of economic economic outcomes. First, we note that administrative remoteness affects the rural employment structure. We find that administrative remoteness reduces the share of village workforce engaged in non-farm activities, potentially through its effect on access to paved roads and other public gooods. The non-farm employment share is 0.7 percentage points lower in treatment villages as compared to control villages. As a percentage of the sample mean, this is equivalent to a 2.8 percent decrease in non-farm employment share in treatment villages. This result is robust, though smaller in magnitude, when we drop state borders. Figure 10 presents this result graphically: there is a drop in non-farm employment share at the district border. A large literature in economics has documented differences in productivity between agriculture and other sectors of the economy (Gollin and Rogerson, 2014). Structural transformation of the economy from agriculture to other economic sectors is key to income growth in developing countries. Our results suggest that administrative remoteness inhibits the structural transformation of the economy away from agriculture.

One explanation for this result could be differences in public sector employment. Greater distance to administrative headquarters could reduce rural residents' access to government jobs and hence drive differences in the employment structure. We find that the share of households that report a government job as the main source of earnings is 0.07 percentage

points lower in treatment villages as compared to control villages. This magnitude is a small proportion of the remoteness effect on the non-farm employment share. Further, the difference in the government employment share is smaller (and statistically indistinguishable from zero) when we drop state borders. This is consistent with recent evidence that documents state-level biases in public sector recruitment in India (Kone et al., 2017). Residents of administratively remote villages can find public sector employment in the neighboring district, but only if the neighboring district belongs to their state.

Second, administrative remoteness has small effects on household income. Average monthly income in treatment villages is approximately 50 rupees (1% of sample mean) lower than in control villages. However, this result is not robust to dropping state borders. Residents of treatment villages may be commuting to the other side of the border in search of employment opportunities. This ensures equilibrium in observed nominal incomes (except where mobility is restricted, such as state borders), even though commuting costs may be higher for administratively remote rural individuals.

Third, along with the income effect, we observe small effects on household assets. For example, the share of households that have a solid wall is 0.9 percentage points lower, the share of households that have a solid roof is 0.4 percentage points lower, and the share of households that own any vehicle is 0.4 percentage points lower in treatment villages. Once again, these magnitudes are fairly small as compared to the outcome mean, and smaller yet when we exclude state borders. The income and asset effects point to the importance of state borders in restricting commuting and labor mobility in rural India.

Finally, administrative remoteness reduces the literacy rate. In treatment villages, the share of literates in the village population is 0.3 percentage points lower than in control villages. This could be driven by the reduced access to government primary schools in treatment villages. However, similar to incomes and assets, when we exclude state borders, the magnitude of the remoteness effect on literacy is smaller. This suggests that even when

villages do not have primary schools, their young residents can cross the border to enroll at the nearest village school in the neighboring district. But this is harder to do when the neighboring district is in another state.

5.4 Robustness

In this section we investigate whether our results are driven by the set of choices we have made while selecting our sample or while specifying the regression discontinuity equation. We find no evidence that that is the case.

One threat to our identification comes from district borders that coincide with rivers. We have assumed that villages on both sides of the border have similar access to market opportunities, but this is not the case when villagers have to cross a river to reach the nearest town. Therefore, we re-estimate our results after dropping all borders that coincide with rivers. In Table A1, we show that administrative remoteness reduces access to public goods provided by the district administration even after restricting our sampleto non-river borders. Additionally, now we cannot reject the null that remoteness does not reduce access to a few federal public goods, such as electricity and national highways. Further, Table A2 shows that remoteness reduces non-farm employment share and worsens incomes, assets and literacy across the set of district borders that do not coincide with rivers.

Another concern is that our results are driven by our RD choices, such as the size of the bandwidth around the district border, the length of the border segments, the specification of the RD polynomial of latitude and longitude, and how the standard errors are clustered. We show here that our results are robust to changing each of those decisions.

First, we use a bandwidth of 6 kilometers around the district border in our preferred specification, which allows us to compare villages located in close geographic proximity while also giving us sufficient power to test our hypotheses. We show here that our results are robust to changing the size of the bandwidth. In Section 4.1, we have already discussed how

the coefficients on distance to towns with population exceeding 10,000, 50,000, 100,000, and 500,000 change when we change the bandwidth from 1 to 9 kilometers around the border. Now, Figure A1 shows how the coefficient on paved roads changes as we change the bandwidth similarly from 1 to 9 kilometers. Except for small bandwidths of 1 and 2 kilometers, where we do not have sufficient power to reject the null, we find that we can always reject the null that administrative remoteness does not reduce access to paved roads. In a similar way, Figure A2 shows how the coefficient on non-farm employment share change as we change the bandwidth. Though estimates using small bandwidths have large standard errors, we find that we can still always reject the null that administrative remoteness reduces the village's share of non-farm employment.

In the interest of brevity, we do not show similar coefficient plots for all other outcomes variables. However, for an intermediate bandwidth of 3 kilometers, we show that administrative remoteness reduces access to public goods provided by the district administration (except for mobile health clinics, for which we cannot reject the null of no effect), but not to public goods provided by other tiers of administration (Table A3). Further, even with a 3 kilometer bandwidth, administrative remoteness reduces the village's non-farm employment share, and reduces household assets and the literacy rate (Table A4).

Second, we divide each border into segments that are 20 kilometers long to ensure comparison between geographically proximate villages. Tables A5 and A6 show that our results are robust to changing the length of the border segments to 40 kilometers.

Third, our preferred RD polynomial is linear in latitude and longitude and estimated separately for each district border, following Gelman and Imbens (2017) and Dell and Olken (2017). We test whether the results are robust to using a quadratic polynomial in latitude and longitude. Tables A7 and A8 show that they are.

Finally, we have clustered our standard errors at the level of district and segment, since that is the unit of treatment in our analysis. We test whether our results are robust to clustering

the standard errors in 10 x 10 kilometer grid cells to allow for spatial correlation in the error term, as done by Burgess et al. (2016). In Tables A9 and A10, we show that standard errors obtained using clustering at the level of grid cells are often smaller than standard errors obtained using clustering at the district-segment level, and our results continue to hold.

6 Mechanisms

Why does distance to administration reduce the availability of public goods? There are several possible reasons, from the perspectives of both the state (the supplier of public goods) and the citizens (the demanders for public goods).

From the perspective of the state, constructing infrastructure can be expensive. These costs may depend on how far the village is from the administrative headquarters. For example, if the materials for constructing roads are sourced in the headquarters, shipment costs to get them to the village can increase with distance. A related reason is corruption: it may be more difficult for administrators to observe expenditures and output quality in more distant locations, allowing contractors to inflate costs. Given a fixed budget and higher costs in more distant locations, administrators may decide to prioritise nearby locations over distant ones.

Even after infrastructure is constructed, it requires regular maintainance. Roads get filled with potholes, canals require desilting. Outside the normal maintainance cycle, there are primarily two ways through which administrators learn about infrastructure quality: one, ground reports from villages, and two, by visiting the villages themselves. When the village is distant from the headquarters, both ground reports and official visits may be less frequent. Even when administrators learn about degradation in quality, it can be more expensive to repair infrastructure due to the shoe-leather cost and corruption channels explained in the previous paragraph.

From the perspective of the citizens, greater distance to administration can weaken their ability to organize and demand public goods from the state. For example, when a village

does not have a paved road, its residents can either wait for the state to build a road in the distant future, or they can organize, petition and lobby with politicians and bureuacrats to persuade them to construct a road. Often, villagers will have to travel frequently to administrative headquarters to articulate these demands. This can be harder to do when the administrative headquarters is a full day's journey away. The demand and supply channels can, in turn, reinforce each other. For example, in the absence of public goods, citizens can disengage from the political process and acquire a more negative view of the state and its ability to deliver them public goods and services. Krishna and Schober (2014) provides descriptive evidence for such a channel in southern Indian villages.

To test some of these mechanisms, we exploit detailed project data from the Pradhan Mantri Gram Sadak Yojna (PMGSY), a national program launched in 2000 to build paved roads in villages all over the country. Until 2015, the program had been used to construct paved access roads in more than 185,000 rural habitations. These roads were built to identical national standards. District administrations were allocated key responsibilities for designing the PMGSY network in their districts, deciding the priority order for construction, seeking bids from contractors and awarding them construction contracts, and supervising the construction process through quality inspectors.

We assemble data on the cost for constructing each road, the time taken to build the road, and road characteristics such as length and type of surface. Additionally, we have data on whether the road quality was audited. We use this data to test how administrative remoteness affects costs and monitoring in national infrastructure projects.

Table 5 presents the results. First, administrative remoteness does not affect the cost of building the roads, or the time taken for the construction. The result holds even when we control for the type of surface of the road constructed under the program. This is relevant since different types of surfaces have different average costs. One explanation for the result could be that more expensive roads are located in treatment villages and these are less likely

to get constructed. However, we note that remote villages are actually more likely to receive a road under the program. Therefore, administrative remoteness does not seem to affect the cost margin, at least not in the PMGSY program.

After construction, though, roads in treatment villages are 2.1 percentage points less likely to get audited through quality inspectors. Even though assignment of roads to audit was random, we find that district administrations are significantly less likely to assign auditors to roads constructed in more distant villages. This suggests administrative remoteness affects monitoring costs.

7 Conclusion

Many countries have committed to providing universal access to public goods and services that are both intrinsically valuable as well as important ingredients for private economic activities. Yet, many people continue to live without these public goods, restricting economic growth and exacerbating inequalities in living standards. A long literature in economics has studied the factors that constrain the state's effectiveness in developing countries, studying its administrative machinery and its citizens' ability to make the administration work in their interest.

However, both the state's officials and its citizens are constrained by the geography of its administration: where the officials are based relative to where its citizens live. In a world of high transport costs and information frictions, this becomes an additional channel through which the state's effectiveness is curtailed.

In this paper we estimate the costs of "administrative remoteness," or the citizens' physical distance from their administration. We use a rich dataset on rural public goods and household economic outcomes covering the universe of Indian villages, and exploit spatial discontinuities in distance to administration by comparing villages located across district borders. Such villages are similar in all respects except in their distance to administration.

We find that administrative remoteness reduces rural access to paved roads, schools, health centers and irrigation facilities, and this in turn worsens rural economic outcomes.

While our data provides us a comprehensive picture of the outcomes of the state's activities, we have much more limited information on the state's inputs into development work. Using implementation data from India's national rural roads program, we find that administrative remoteness does not affect the cost or duration of road construction in rural habitations. However, roads constructed in more remote villages are less likely to be audited by national quality monitors. Reduced monitoring can lead to the construction of worse roads in rural areas (Olken, 2007), squandering fiscal resources and denying villagers the opportunity to access external markets. However, the relationship between the state and its citizens is correlative. When denied critical public inputs, how do citizens organize to demand public goods from the state? Our data does not allow us to observe and measure citizen actions in response to the underprovision of public goods. Such questions should be the subject of future research.

Few questions have received more attention from social scientists than the constraints on state effectiveness and how to make the states work better in the interest of their citizens. Our results suggest that the spatial organization of public administration is an important barrier to state effectiveness. Reducing the distance between the citizens and the state can help to improve the state's ability to provide and monitor public goods provision in rural areas and reduce the spatial inequality in living standards observed across many developing countries today. However, policies that reduce distance, such as redrawing administrative borders or changing the location of administrative headquarters or creating smaller administrative units, may also pose additional fiscal costs. Developing countries would have to evaluate whether the additional benefits in terms of expanded rural opportunities and reduced spatial disparities are worth the cost.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *arXiv preprint arXiv:1710.02926*, 2017.

Adamopoulos, Tasso, "Transportation Costs, Agricultural Productivity, And Cross-Country Income Differences," *International Economic Review*, 2011, *52* (2), 489–521.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, 2003, *8* (2), 155–194.

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Anderson, Siwan, Patrick Francois, and Ashok Kotwal, "Clientelism in Indian villages," *The American Economic Review*, 2015, *105* (6), 1780–1816.

Asher, Sam and Paul Novosad, "The Employment Effects of Road Construction in Rural India," 2017.

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack, "No margin, no mission? A field experiment on incentives for public service delivery," *Journal of Public Economics*, 2014, *120*, 1–17.

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.

Banerjee, A, R Hanna, and S Mullainathan, "Corruption, the Handbook of Organizational Economics," 2013.

Banerjee, Abhijit and Lakshmi Iyer, "History, institutions, and economic performance: the legacy of colonial land tenure systems in India," *The American Economic Review*, 2005, *95* (4), 1190–1213.

——, ——, and Rohini Somanathan, "History, social divisions, and public goods in rural India," *Journal of the European Economic Association*, 2005, *3* (2-3), 639–647.

——, ——, and ——, "Public action for public goods," *Handbook of Development Economics*, 2007, *4*, 3117–3154.

Barankay, Iwan and Ben Lockwood, "Decentralization and the productive efficiency of government: Evidence from Swiss cantons," *Journal of Public Economics*, 2007, *91* (5), 1197–1218.

Bardhan, Pranab, "The Political Economy of Development in India," 1984.

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

Burgess, Robin, Francisco Costa, and Benjamin Olken, "The Power of the State: National Borders and the Deforestation of the Amazon," 2016.

——, Remi Jedwab, Edward Miguel, Ameet Morjaria et al., "The value of democracy: evidence from road building in Kenya," *The American Economic Review*, 2015, *105* (6), 1817–1851.

Callen, Michael Joseph, Saad Gulzar, Syed Ali Hasanain, and Muhammad Yasir Khan, "The political economy of public employee absence: Experimental evidence from Pakistan," 2013.

Caselli, Francesco, "Accounting for cross-country income differences," *Handbook of Economic Growth*, 2005, *1*, 679–741.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

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

1863–1903.

—— **and Benjamin A Olken**, "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java," 2017.

—— **, Nathaniel Lane, and Pablo Querubin**, "The Historical State, Local Collective Action, and Economic Development in Vietnam," Technical Report, National Bureau of Economic Research 2017.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *The American Economic Review*, 2001, *91* (4), 795–813.

Fafchamps, Marcel and Jackline Wahba, "Child labor, urban proximity, and household composition," *Journal of Development Economics*, 2006, *79* (2), 374–397.

Faguet, Jean-Paul, "Does decentralization increase government responsiveness to local needs?: Evidence from Bolivia," *Journal of Public Economics*, 2004, *88* (3), 867–893.

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

Finan, Frederico, Benjamin A Olken, and Rohini Pande, "The personnel economics of the state," Technical Report, National Bureau of Economic Research 2015.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky, "School decentralization: Helping the good get better, but leaving the poor behind," *Journal of Public Economics*, 2008, *92* (10), 2106–2120.

Gelman, Andrew and Guido Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 2017.

Gollin, Douglas and Richard Rogerson, "Productivity, transport costs and subsistence agriculture," *Journal of Development Economics*, 2014, *107*, 38–48.

—— **, David Lagakos, and Michael E Waugh**, "Agricultural productivity differences across countries," *The American Economic Review*, 2014, *104* (5), 165–170.

Green, Elliott, "Patronage, district creation, and reform in Uganda," *Studies in comparative international development*, 2010, *45* (1), 83–103.

Grossman, Guy and Janet I Lewis, "Administrative unit proliferation," *American Political Science Review*, 2014, *108* (1), 196–217.

—— **, Jan H Pierskalla, and Emma Boswell Dean**, "Government fragmentation and public goods provision," *The Journal of Politics*, 2017, *79* (3), 823–840.

Hanna, Rema and Shing-Yi Wang, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, 2017, *9* (3), 262–290.

Hassan, Mai, "A state of change: district creation in Kenya after the beginning of multi-party elections," *Political Research Quarterly*, 2016, *69* (3), 510–521.

Iyer, Lakshmi, "Direct versus indirect colonial rule in India: Long-term consequences," *The Review of Economics and Statistics*, 2010, *92* (4), 693–713.

Kis-Katos, Krisztina and Bambang Suharnoko Sjahrir, "The impact of fiscal and political decentralization on local public investment in Indonesia," *Journal of Comparative Economics*, 2017, *45* (2), 344–365.

Kone, Zovanga, Maggie Liu, Aaditya Mattoo, Caglar Ozden, and Siddharth Sharma, "Internal Borders and Migration in India," 2017.

Kothari, Rajni, *Politics in India*, Orient Blackswan, 1970.

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

Mansuri, Ghazala and Vijayendra Rao, *Localizing development: does participation work?*, World Bank Publications, 2013.

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

Mookherjee, Dilip, "Political decentralization," *Annual Review of Economics*, 2015, *7* (1), 231–249.

Munshi, Kaivan and Mark Rosenzweig, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *The American Economic Review*, 2016, *106* (1), 46–98.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher performance pay: Experimental evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

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

Oates, Wallace E, "Fiscal Federalism," *Harcourt Brace Jovanovich*, 1972.

Olken, Benjamin A, "Monitoring corruption: evidence from a field experiment in Indonesia," *Journal of Political Economy*, 2007, *115* (2), 200–249.

Pierskalla, Jan H, "Splitting the difference? The politics of district creation in Indonesia," *Comparative Politics*, 2016, *48* (2), 249–268.

Rasul, Imran and Daniel Rogger, "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service," *The Economic Journal*, forthcoming.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of Monetary Economics*, 2008, *55* (2), 234–250.

Second Administrative Reforms Commission, "Fifteenth report: State and District Administration," Technical Report, Second Administrative Reforms Commission 2009.

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 Hindu, "What is the special category status?," August 2016.

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

Table 1: Changes at the district border

	β_1	(se)	Sample mean	N
First stage				
Distance to District HQ	15.579	(.223)***	41.87 km	180,331
Distance to towns				
Pop > 10,000	0.047	(0.073)	15.88 km	180,331
Pop > 50,000	0.436	(0.064)***	29.14 km	180,331
Pop > 100,000	0.776	(0.064)***	40.66 km	180,331
Pop > 500,000	-0.031	(0.038)	90.81 km	180,331
Demographic variables				
Population 2011	-3.188	(10.855)	1,397	180,331
Scheduled Caste share	0.027	(0.149)	18.89%	180,331
Scheduled Tribe share	0.195	(0.197)	16.59%	180,331
Geographic variables				
Altitude	0.510	(0.488)	242.6 meters	180,331
Distance to river	0.155	(0.051)***	52.09 kms	180,331
Bandwidth: 6 km				
Geographic polynomial: Linear in latitude and longitude				

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

Notes: This table presents regression coefficients from estimating Equation 1. The left hand side variables are listed in Column 1. Column 2 presents the point estimate for β_1, the coefficient on the binary treatment variable (a village is considered treated if it is located on the more administratively remote side of the district border). Column 3 presents the standard errors, clustered at the level of district-segment. Column 4 presents the sample mean for the left hand side variables, and Column 5 has the number of observations. The regression in the first row includes the full set of controls. Every other regression includes the full set of controls except for the variable on the left hand side. All regressions include a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad, and district and segment fixed effects.

Table 2: Effects on paved road access

	β_1	(se)	Sample mean	N	Cluster	Bandwidth	RD Polynomial
Paved road access	-1.232	(.409)***	69.56%	180,264	District-segment	6 km	Linear Lat-Long
		(.449)***			Spatial	6 km	Linear Lat-Long
Different RD polynomials							
Paved road access	-1.137	(.575)**	69.56%	180,264	District-segment	6 km	Distance to border
Paved road access	-1.009	(.600)*	69.56%	180,264	District-segment	6 km	Both
Paved road access	-0.950	(.377)**	69.56%	180,264	District-segment	6 km	Quadratic Lat-Long
Drop riverine borders							
Paved road access	-1.583	(.437)***	69.16%	157,395	District-segment	6 km	Linear Lat-Long
Drop state borders							
Paved road access	-1.654	(.598)***	70.33%	141,621	District-segment	6 km	Linear Lat-Long
Restrict bandwidth							
Paved road access	-1.490	(.743)***	70.67%	76,083	District-segment	3 km	Linear Lat-Long
Probability of new roads							
Paved road access	-1.278	(.692)*	71.45%	82,767	District-segment	6 km	Linear Lat-Long
Intensity of treatment							
Paved road access (25th-75th pct)	-1.843	(.447)***	69.56%	180,264	District-segment	6 km	Linear Lat-Long

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to paved roads. Paved road access is measured using a binary variable that takes the value 1 if the village is recorded as having an "all-weather road" according to the Village Directory of Population Census 2011. Column 2 presents the point estimate for β_1, the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. Column 3 presents the standard errors. In all rows except Row 3, these are clustered at the district-segment level. In Row 3, we cluster standard errors allowing for spatial correlation, using 10x10 km grid cells. Column 4 and 5 present the sample mean and sample size respectively. Column 6 states the bandwidth used around the district border. Except for Row 7, we employ a 6 km bandwidth around the district border. In Row 7, we employ a bandwidth of 3 km. Finally, Column 7 states the RD polynomial used. Except for Rows 3-5, this is a linear polynomial in latitude and longitude, estimated separately for each district border. In Row 3, the RD polynomial is a linear polynomial in distance to district border (inversed for the control villages). In Row 4, the RD polynomial is a linear polynomial in distance to district border (inversed for control villages), and latitude and longitude. In Row 5, the polynomial is a quadratic polynomial in latitude and longitude. Row 8, the second last row in the table, restricts the sample to villages that did not have a paved road in 2001. The left hand side variable is the probability that the village did not have a paved road in 2001 but did have one in 2011. The last row of the table, Row 9, presents estimates from Equation 2, where we regress paved road access on distance to district headquarters, rather than the binary treatment variable. The magnitude of β_1 in the last row is the decrease in the probability that a village has a paved road when we go from the 25th percentile (25 kilometers) to the 75th percentile (53 kilometers) of the distance to district HQ distribution in our sample. All regressions include district and border segment fixed effects.

Table 3: Rural access to different public goods

	All district borders			Without state borders		
	β_1 (se)	Sample mean	N	β_1 (se)	Sample mean	N
Villages with:						
Amenities provided by districts:						
Paved roads	-1.232	69.56%	180,264	-1.654	70.33%	141,621
	(0.409)***			(0.598)***		
Govt primary schools	-0.947	83.10%	179,904	-1.097	83.05%	141,364
	(0.257)***			(0.283)***		
Mobile health clinic	-0.310	1.84%	180,135	-0.208	1.57%	141,547
	(0.092)***			(0.094)**		
Share of land irrigated	-0.708	44.76%	180,299	-0.286	46.25%	141,692
	(0.231)***			(0.262)		
Treated water	-1.190	21.82%	178,606	-1.121	22.03%	140,252
	(0.279)***			(0.321)***		
Amenities provided by federal govt						
Post office	0.021	9.99%	178,608	0.161	10.31%	140,251
	(0.190)			(0.199)		
Polling station	-0.325	68.21%	178,595	-0.329	70.05%	140,239
	(0.295)			(0.309)		
National highway	-0.259	5.02%	178,609	-0.211	5.28%	140,252
	(0.172)			(0.198)		
Electricity	-0.542	55.26%	180,299	0.066	57.28%	141,692
	(0.368)			(0.419)		
Amenities provided by village councils						
Gravel road	0.273	85.57%	178,569	0.332	85.08%	140,217
	(0.293)			(0.339)		
Community toilet complex	-0.019	4.95%	178,609	-0.066	2.38%	140,252
	(0.095)			(0.109)		
Community well	-0.425	56.69%	178,602	-0.458	54.69%	140,252
	(0.370)			(0.429)		
Tubewell	-0.261	50.49%	178,595	-0.152	52.83%	140,252
	(0.415)			(0.471)		

Bandwidth: 6 km
Geographic polynomial: Linear in latitude and longitude
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table 4: Effects on rural incomes, assets and employment

	All district borders			Without state borders		
	β_1 (se)	Sample mean	N	β_1 (se)	Sample mean	N
Employment structure						
Non-farm share	-0.733	26.58%	178,457	-0.577	26.89%	140,279
	(0.208)***			(0.220)***		
Govt salary share	-0.076	3.75%	171,425	-0.026	3.88%	134,777
	(0.043)*			(0.049)		
Enterprise share	-0.035	1.12%	171,423	-0.078	1.11%	134,775
	(0.055)			(0.045)*		
Incomes						
Average monthly income	-50.468	4,893	171,422	-18.566	4,946	134,774
	(18.844)***			(20.841)		
Household assets						
Solid wall share	-0.871	53.22%	171,422	-0.675	54.68%	134,774
	(0.229)***			(0.251)***		
Solid roof share	-0.384	44.36%	171,422	-0.100	45.77%	134,774
	(0.212)*			(0.233)		
Vehicle ownership share	-0.365	19.32%	171,425	-0.273	20.03%	134,777
	(0.157)**			(0.168)		
Literacy						
Literacy rate	-0.261	55.99%	178,465	-0.156	57.34%	140,269
	(0.087)***			(0.094)*		
Bandwidth: 6 km						
Geographic polynomial: Linear in latitude and longitude						

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

Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

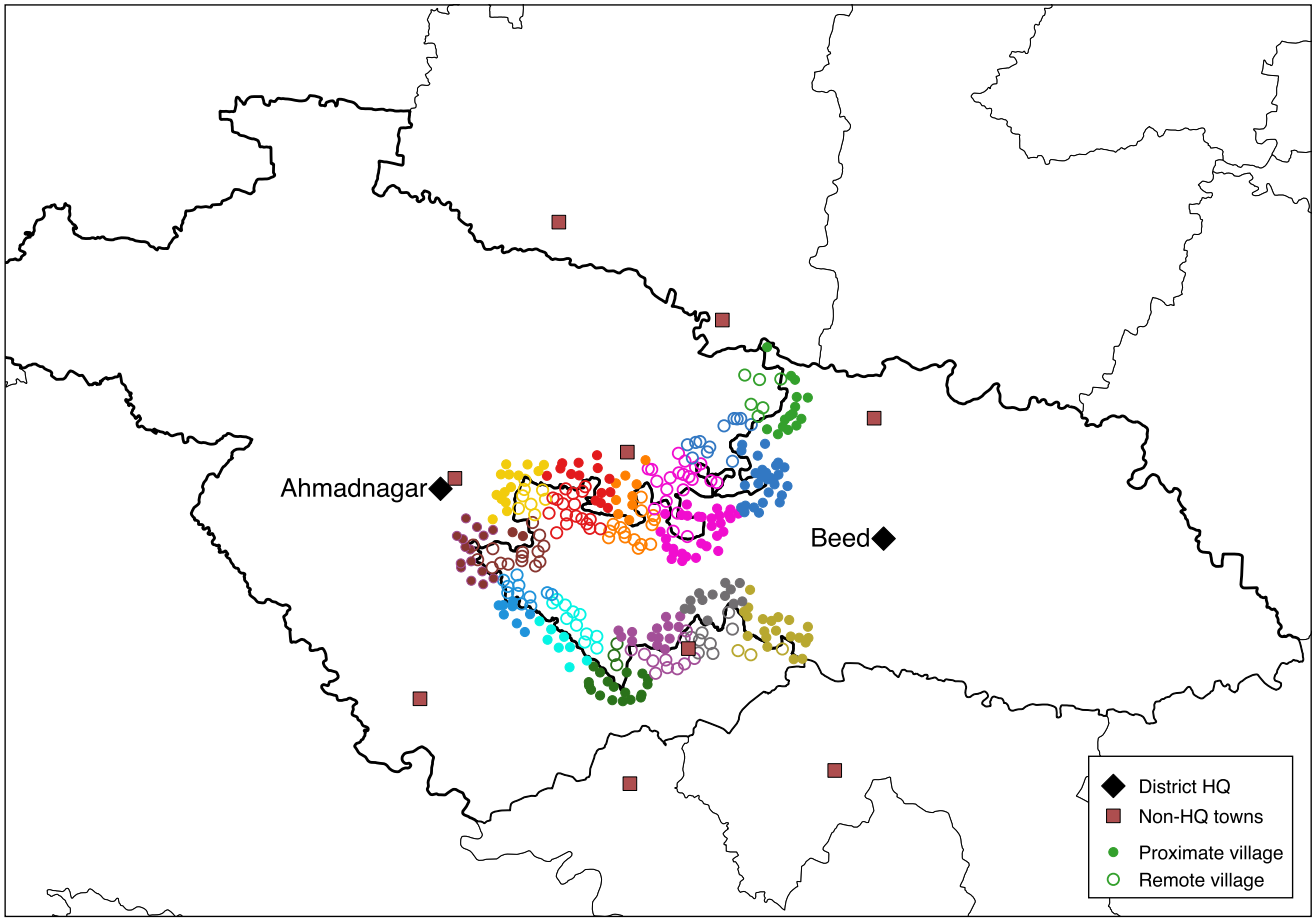
Table 5: PMGSY Program Information

	β_1	(se)	Sample mean	N
Placement				
Received PMGSY road	0.655	(.225)**	16.73%	180,299
Project outcomes				
Estimated cost per km	-0.007	(.020)	INR 3.21 million	27,147
Actual cost per km	-0.003	(0.020)	INR 2.49 million	19,858
Estimated duration per km	-1.271	(2.351)	145.8 days	24,052
Actual duration per km	-1.684	(4.951)	270.3 days	22,489
Controlling for surface type				
Estimated cost per km	0.012	(0.022)	INR 3.21 million	22,273
Actual cost per km	-0.012	(0.024)	INR 2.49 million	15,369
Estimated duration per km	-2.381	(2.649)	145.8 days	19,405
Actual duration per km	-6.667	(5.974)	270.3 days	17,925
Monitoring inspection				
Probability of quality audit	-2.192	(0.834)***	23.09%	30,170
Bandwidth: 6 km				
Geographic polynomial: Linear in latitude and longitude				

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

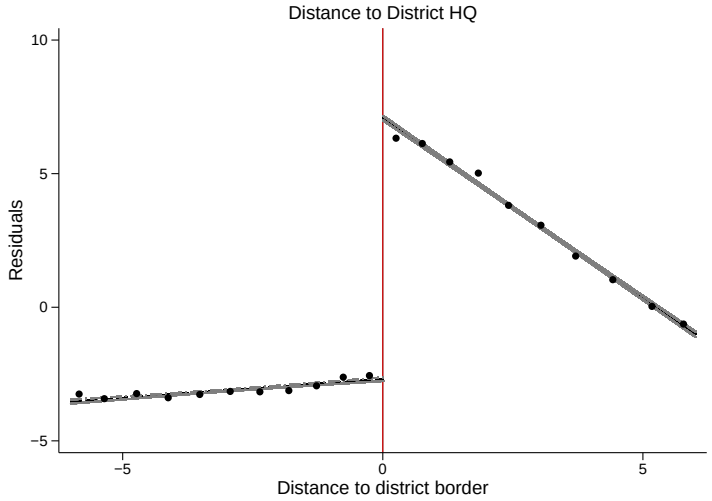
Notes: This table presents regression coefficients from estimating Equation 1 for PMGSY program outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. In some specifications, we also control for type of road surface. Cost figures are in millions of Indian rupees and duration figures are in number of days. All regressions include district and border segment fixed effects. We control for geographic location using a linear polynomial in latitude and longitude. Standard errors are clustered at the district-segment level.

Figure 1: Illustration of Empirical Strategy



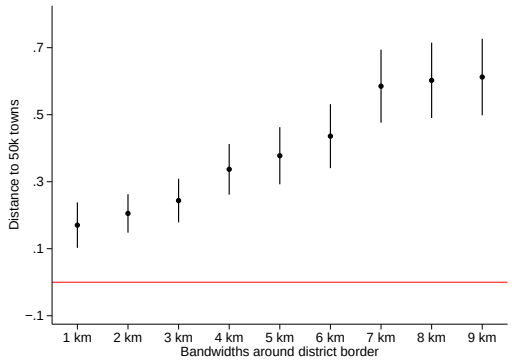
Notes: Example of our empirical strategy, showing Beed and Ahmadnagar districts in Maharashtra, with their district headquarters, border villages, and towns with 2011 population greater than 10,000. Different colored dots represent different segments of the district border. Filled circles are villages located on the side of the border that is closer to its administrative headquarters, hollow circles are villages located on the side of the border that is more distant to its administrative headquarters.

Figure 2: RD plot: first stage



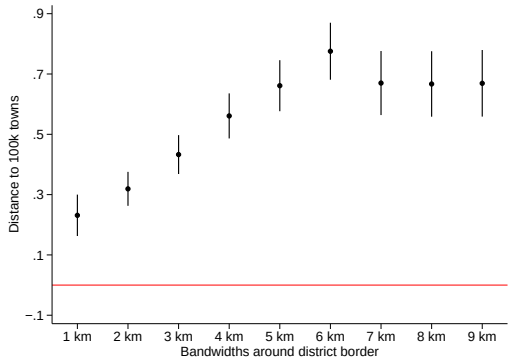
Notes: The figure plots residualized distance to district headquarters, after controlling for all balance variables: distance to towns with 2011 population greater than 10,000, 50,000, 100,000, and 500,000, percent of village population that belongs to Scheduled Castes and Scheduled Tribes, mean elevation, distance to the nearest major river, as well as a linear polynomial in latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero represent administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages. Each point represents approximately 6,000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of the district border. It excludes villages in special category states, villages along riverine district borders, and villages along district borders with large elevation changes.

Figure 3: Distance to towns with population greater than 50,000, using different bandwidths



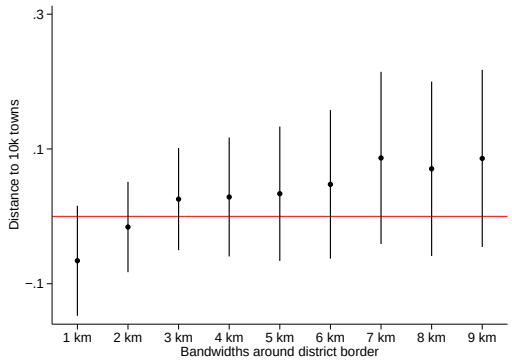
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 50,000 on all right hand side variables in Equation 1, except itself, for different bandwidths around the district border.

Figure 4: Distance to towns with population greater than 100,000, using different bandwidths



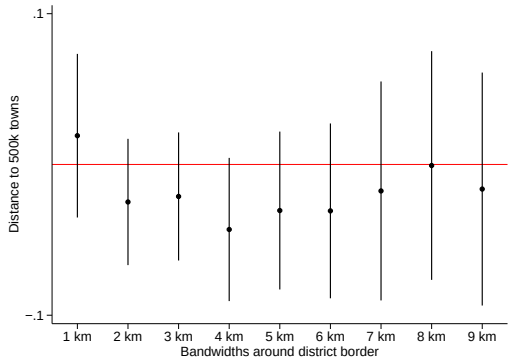
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 100,000 on all right hand side variables in Equation 1, except itself, for different bandwidths around the district border.

Figure 5: Distance to towns with population greater than 10,000, using different bandwidths



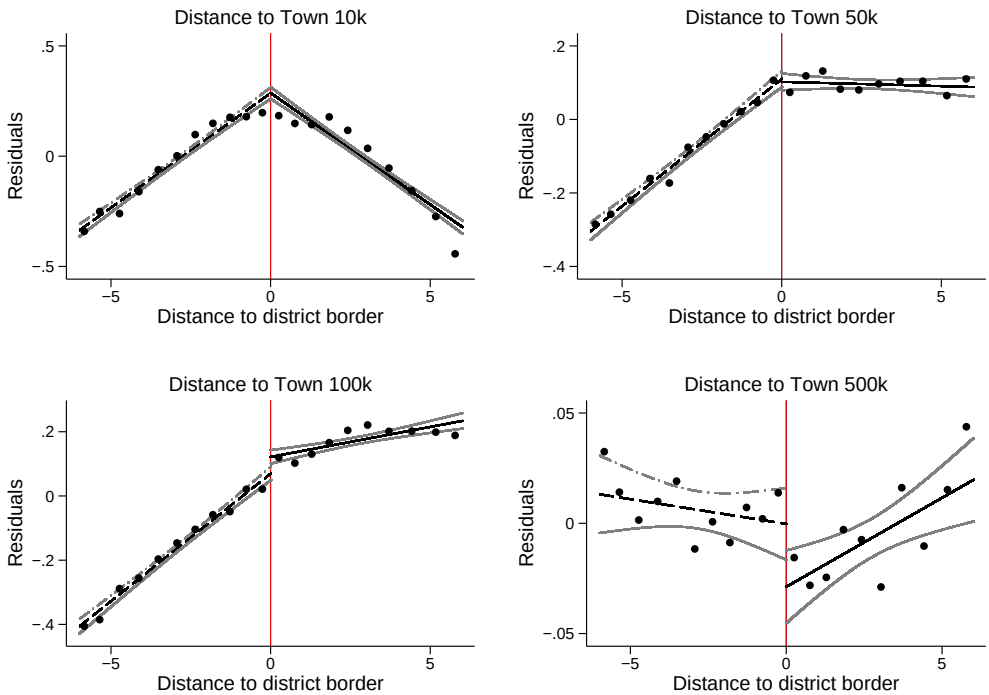
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 10,000 on all right hand side variables in Equation 1, except itself, for different bandwidths around the district border.

Figure 6: Distance to towns with population greater than 500,000, using different bandwidths



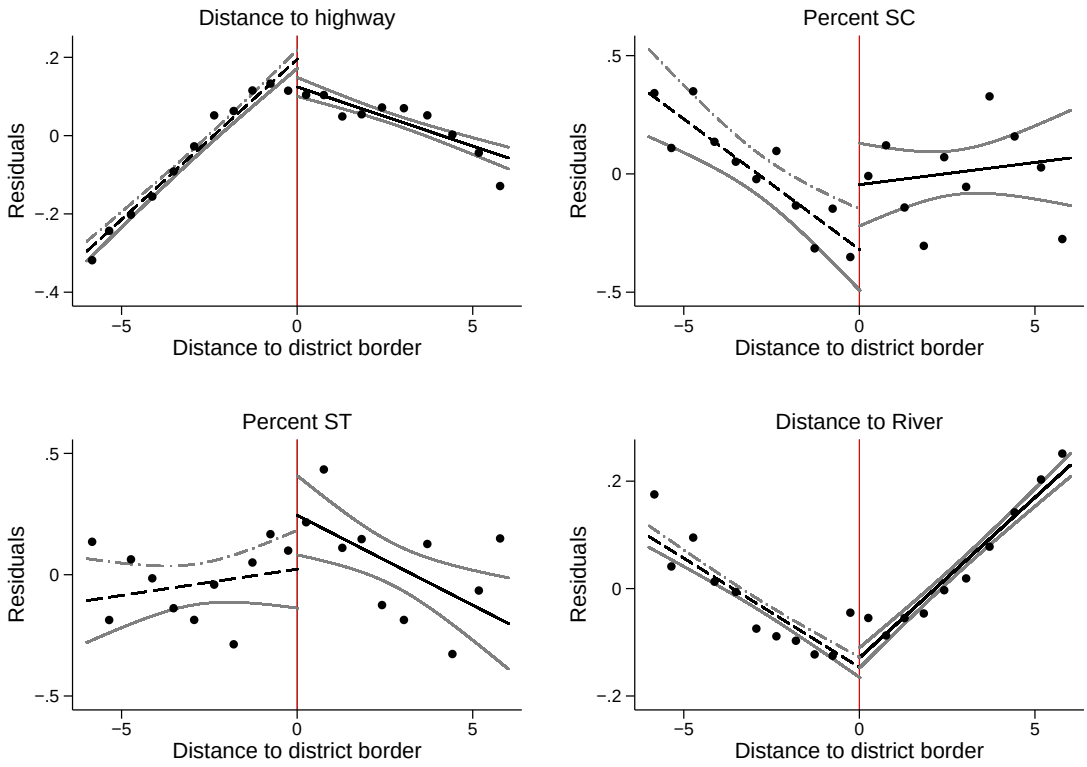
Notes: The figure plots the coefficient β_1 when we regress distance to town with population greater than 500,000 on all right hand side variables in Equation 1, except itself, for different bandwidths around the district border.

Figure 7: RD plot: balance on distance variables



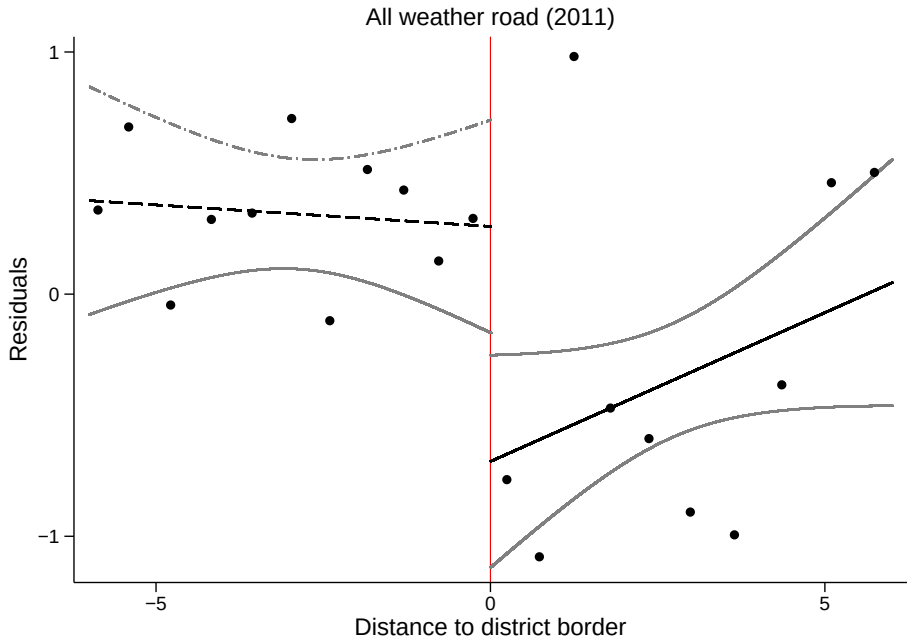
Notes: The figure plots residualized values of distance-to-town balance variables, controlling for all other balance variables other than the one on the left-hand side, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of an intra-state district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 8: RD plot: balance on other variables



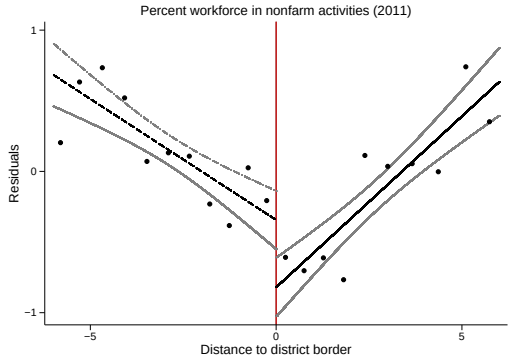
Notes: The figure plots residualized values of balance variables other than distance to towns, controlling for all balance variables other than the one on the left-hand side, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of an intra-state district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 9: RD plot: access to roads



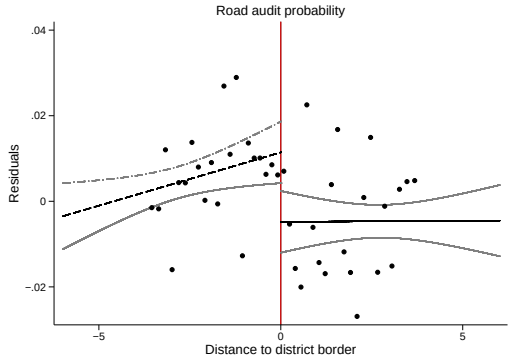
Notes: The figure plots the probability that a village has an all-weather (paved) road according to the Village Directory of Population Census 2011 controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of a district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 10: RD plot: rural non-farm employment



Notes: The figure plots the residualized share of village workforce engaged in non-agricultural activities according to the 2011 Population Census, controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 6,000 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of a district border. It excludes villages in special category states, villages that lie along riverine district borders, and villages that lie along district borders with large elevation changes.

Figure 11: RD plot: road monitoring



Notes: The figure plots the probability that, conditional on receiving a road under the PMGSY program, the village road was audited by a national quality monitoring inspector, controlling for all balance variables, as well as a linear polynomial of latitude and longitude that is estimated separately for each district border (district-district dyad). Points to the right of zero are administratively remote villages that are more distant to their district HQ than neighbouring villages on the other side of the border, while points to the left of zero are administratively proximate villages that are closer to their district HQ than neighbouring villages on the other side of the district border. Each point represents approximately 1,750 observations. A linear fit is generated separately for each side of 0, with 95% confidence intervals displayed. The sample consists of villages within 6 kilometers of a district border. It excludes villages in special category states, and villages that lie along district borders with large elevation changes.

A Appendix: Additional figures and tables

Table A1: Rural access to public goods, after dropping river borders

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts			
Paved roads	-1.583	69.16%	157,395
	(.437)***		
Govt primary schools	-0.840	82.97%	157,082
	(0.267)***		
Mobile health clinic	-0.376	1.88%	157,339
	(0.098)***		
Share of land irrigated	-0.688	44.77%	157,450
	(0.248)***		
Treated water	-0.908	22.05%	155,928
	(0.295)***		
Amenities provided by central govt			
Post office	-0.017	9.86%	155,927
	(0.193)		
Polling station	-0.337	67.9%	155,914
	(0.301)		
National highway	-0.591	4.86%	155,928
	(0.180)***		
Electricity	-0.677	54.81%	157,450
	(0.396)*		
Amenities provided by village councils			
Gravel road	0.292	85.62%	155,888
	(0.323)		
Community toilet complex	0.002	2.25%	155,928
	(0.101)		
Community well	-0.041	56.93%	155,928
	(0.395)		
Tubewell	-0.276	50.37%	155,928
	(0.445)		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A2: Rural incomes, assets, employment, after dropping river borders

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.852	25.99%	155,754
	(0.175)***		
Govt salary share	-0.115	3.68%	149,294
	(0.035)***		
Enterprise share	0.031	1.07%	149,292
	(0.048)		
Incomes			
Average monthly income	-56.284	4,872	149,291
	(15.820)***		
Household assets			
Solid wall share	-0.970	52.81%	149,292
	(0.192)***		
Solid roof share	-0.576	44.03%	149,292
	(0.177)***		
Vehicle ownership share	-0.538	19.28%	149,294
	(0.131)***		
Literacy			
Literacy rate	-0.265	55.48%	155,928
	(0.075)***		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A3: Rural access to public goods, using 3 km bandwidth

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts			
Paved roads	-1.490	70.67%	96,539
	(0.743)***		
Govt primary schools	-1.047	83.30%	96,539
	(0.299)***		
Mobile health clinic	-0.149	1.80%	96,658
	(0.108)		
Share of land irrigated	-0.597	44.85%	96,729
	(0.270)**		
Treated water	-1.224	21.73%	95,846
	(0.330)***		
Amenities provided by federal govt			
Post office	0.097	9.88%	95,845
	(0.233)		
Polling station	-0.036	68.26%	95,837
	(0.336)		
National highway	-0.047	4.88%	95,846
	(0.216)		
Electricity	-0.422	55.00%	96,729
	(0.434)		
Amenities provided by village councils			
Gravel road	0.540	85.71%	95,830
	(0.361)		
Community toilet complex	-0.190	2.21%	95,846
	(0.120)		
Community well	-0.407	56.57%	95,846
	(0.427)		
Tubewell	0.204	50.17%	95,846
	(0.497)		
Bandwidth: 3 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A4: Rural incomes, assets, employment, using 3 km bandwidth

	β_1 (se)	Sample Mean	N
Employment structure			
Non-farm share	-0.952	26.07%	95,640
	(0.191)***		
Govt salary share	-0.069	3.62%	91,808
	(0.040)*		
Enterprise share	-0.044	1.13%	91,806
	(0.066)		
Incomes			
Average monthly income	-28.944	4,869	91,805
	(17.834)		
Household assets			
Solid wall share	-0.742	52.8%	91,807
	(0.214)***		
Solid roof share	-0.448	44.06%	91,807
	(0.189)**		
Vehicle ownership share	-0.159	19.3%	91,808
	(0.137)		
Literacy			
Literacy rate	-0.253	55.71%	95,846
	(0.080)***		
Bandwidth: 3 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A5: Rural access to public goods, using 40 kilometer segments

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by district			
Paved roads	-1.204	69.57%	172,955
	(0.438)***		
Govt primary schools	-0.993	83.13%	172,647
	(0.274)***		
Mobile health clinic	-0.152	1.94%	172,888
	(0.113)		
Share of land irrigated	-0.334	44.33%	173,053
	(0.256)		
Treated water	-0.868	23.02%	171,342
	(0.324)***		
Amenities provided by central govt			
Post office	-0.027	9.27%	171,341
	(0.189)		
Polling station	0.137	68.21%	171,328
	(0.305)		
National highway	0.066	4.94%	171,342
	(0.176)		
Electricity	-0.403	56.85%	173,053
	(0.464)		
Amenities provided by village councils			
Gravel road	-0.111	85.53%	171,302
	(0.314)		
Community toilet complex	-0.050	2.23%	171,342
	(0.104)		
Community well	-0.323	55.16%	171,342
	(0.400)		
Tubewell	0.032	51.36%	171,342
	(0.464)		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A6: Rural incomes, assets, employment, using 40 kilometer segments

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.338	26.60%	171,252
	(0.185)*		
Govt salary share	-0.051	3.70%	164,710
	(0.040)		
Enterprise share	-0.038	1.16%	164,708
	(0.040)		
Incomes			
Average monthly income	-56.014	4,869	164,707
	(15.811)***		
Household assets			
Solid wall share	-0.826	53.63%	164,706
	(0.209)***		
Solid roof share	-0.193	44.70%	164,706
	(0.197)		
Vehicle ownership share	-0.349	19.78%	164,710
	(0.133)***		
Literacy			
Literacy rate	-0.118	56.23%	171,342
	(0.077)		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A7: Rural access to public goods, using quadratic polynomials

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by district			
Paved roads	-0.950	69.56%	180,264
	(.377)**		
Govt primary schools	-0.874	83.10%	179,904
	(0.226)***		
Mobile health clinic	-0.231	1.85%	180,135
	(0.084)***		
Share of land irrigated	-0.573	44.76%	180,299
	(0.216)***		
Treated water	-1.126	21.82%	178,609
	(0.261)***		
Amenities provided by central govt			
Post office	0.047	9.99%	178,608
	(0.161)		
Polling station	-0.035	68.21%	178,595
	(0.252)		
National highway	-0.136	5.02%	178,609
	(0.153)		
Electricity	-1.020	55.26%	180,299
	(0.355)***		
Amenities provided by village councils			
Gravel road	0.081	85.57%	178,569
	(0.277)		
Community toilet complex	0.001	2.18%	178,609
	(0.083)		
Community well	-0.497	56.69%	178,609
	(0.347)		
Tubewell	-0.115	50.49%	178,609
	(0.387)		
Bandwidth: 6 km			
Geographic polynomial: Quadratic in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A8: Rural incomes, assets, employment, using quadratic polynomials

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-1.094	26.58%	178,425
	(0.157)***		
Govt salary share	-0.075	3.75%	171,388
	(0.033)**		
Enterprise share	-0.073	1.12%	171,386
	(0.038)*		
Incomes			
Average monthly income	-28.945	4,893	171,385
	(13.545)***		
Household assets			
Solid wall share	-0.801	53.22%	171,384
	(0.169)***		
Solid roof share	-0.177	44.36%	171,384
	(0.162)		
Vehicle ownership share	-0.424	19.32%	171,388
	(0.110)***		
Literacy			
Literacy rate	-0.239	55.99%	178,609
	(0.066)***		
Bandwidth: 6 km			
Geographic polynomial: Quadratic in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered at the district-segment level.

Table A9: Rural access to public goods, using spatially clustered standard errors

	β_1 (se)	Sample mean	N
Villages with:			
Amenities provided by districts:			
Paved roads	-1.232	69.56%	180,264
	(0.449)***		
Govt primary schools	-0.929	83.10%	179,904
	(0.275)***		
Mobile health clinic	-0.312	1.84%	180,135
	(0.098)***		
Share of land irrigated	-0.710	44.76%	180,299
	(0.250)***		
Treated water	-1.180	21.82%	178,609
	(0.299)***		
Amenities provided by federal govt			
Post office	0.042	9.99%	178,608
	(0.213)		
Polling station	-0.295	68.21%	178,595
	(0.322)		
National highway	-0.255	5.02%	178,609
	(0.199)		
Electricity	-0.535	55.26%	180,299
	(0.402)		
Amenities provided by village councils			
Gravel road	0.277	85.57%	178,569
	(0.317)		
Community toilet complex	-0.017	4.95%	178,609
	(0.112)		
Community well	-0.419	56.69%	178,609
	(0.395)		
Tubewell	-0.256	50.49%	178,609
	(0.457)		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

Notes: This table presents regression coefficients from estimating Equation 1 for rural access to different public goods. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. All left hand side variables, except for "share of land irrigated", are binary variables that take the value 1 if the Village Directory of Population Census 2011 reports that the said public good is available in the village, and 0 otherwise. For "share of land irrigated", we divide the total land area (in hectares) that has assured irrigation for at least one season in the year by the total area (in hectares) of cultivated land in the village. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered using 10x10 km grid cells to allow for spatial correlation in standard errors.

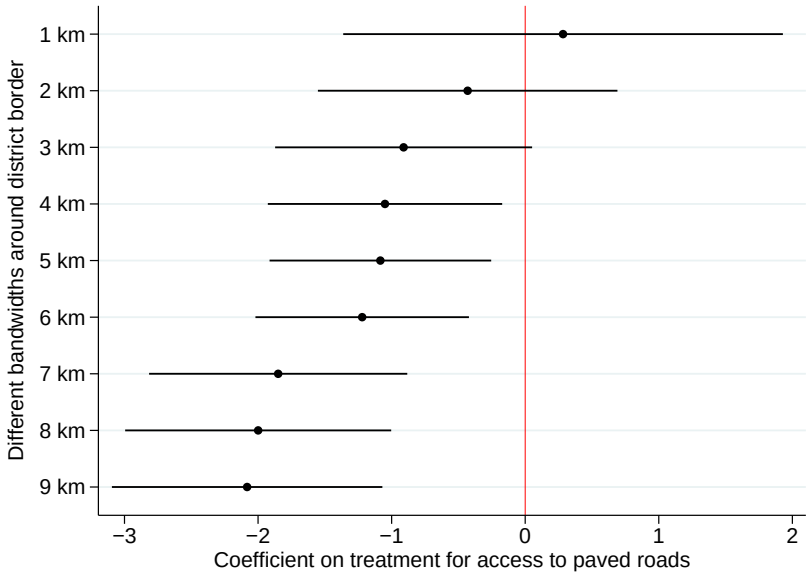
Table A10: Rural incomes, assets, employment, using spatially clustered standard errors

	β_1 (se)	Sample mean	N
Employment structure			
Non-farm share	-0.729	26.58%	178,425
	(0.182)***		
Govt salary share	-0.076	3.75%	171,388
	(0.038)**		
Enterprise share	-0.035	1.12%	171,386
	(0.051)		
Incomes			
Average monthly income	-50.457	4,893	171,385
	(17.323)***		
Household assets			
Solid wall share	-0.871	53.22%	171,384
	(0.197)***		
Solid roof share	-0.384	44.36%	171,384
	(0.176)**		
Vehicle ownership share	-0.365	19.32%	171,388
	(0.134)***		
Literacy			
Literacy rate	-0.261	55.99%	178,609
	(0.077)***		
Bandwidth: 6 km			
Geographic polynomial: Linear in latitude and longitude			

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

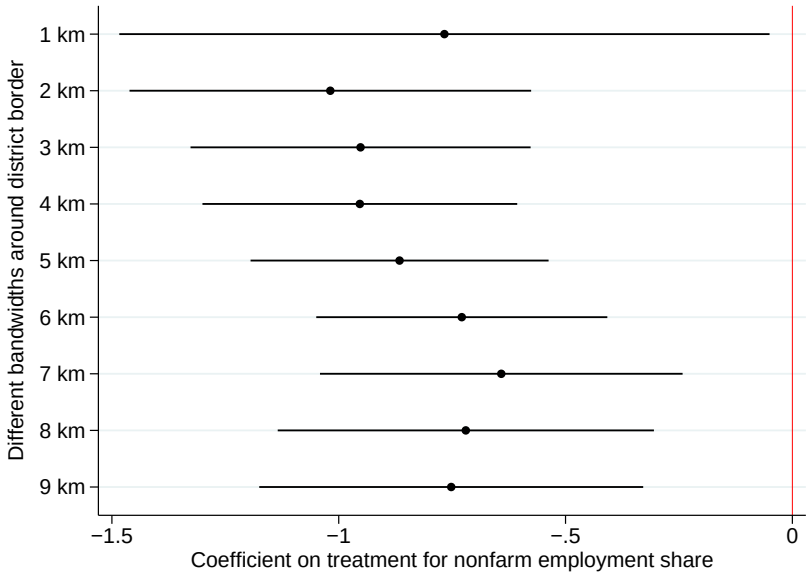
Notes: This table presents regression coefficients from estimating Equation 1 for different rural economic outcomes. β_1 is the coefficient on the binary treatment variable, $RemoteSide_{d,s}$, that takes the value 1 for villages located on the side of the district border that is more distant to district headquarters, and 0 otherwise. "Nonfarm share" is the share of main workers in the village that do not report their main occupation as cultivators or agricultural laborers, as per the Primary Census Abstract of Population Census 2011. "Govt salary share" is the share of households in the village that report a government salary as their main source of income in the Socioeconomic and Caste Census 2012 (SECC 2012). "Enterprise share" is the share of households in the village that report income from an enterprise as the main source ofincome in SECC 2012. "Average monthly income" is calculated as explained in Section 3.2. "Solid wall share", "solid roof share", and "vehicle ownership share" are, respectively, the share of households that have a solid wall, the share of households that have a solid roof, and the share of households that own at least 1 vehicle as per SECC 2012. Finally, "literacy rate" is the share of village population that is literate according to the Primary Census Abstract of the 2011 Population Census. All regressions include district and border segment fixed effects. The RD polynomial is a linear polynomial in latitude and longitude, estimated separately for each border, i.e. each district-district dyad. Standard errors are clustered using 10x10 km grid cells to allow for spatial correlation in standard errors.

Figure A1: Differences in paved road access, using different bandwidths



Notes: This figure plots the coefficient β_1 obtained from estimating Equation 1 for village access to paved roads, using a number of different bandwidths around district border.

Figure A2: Differences in non-farm employment share, using different bandwidths



Notes: This figure plots the coefficient β_1 obtained from estimating Equation 1 for village nonfarm employment share, using a number of different bandwidths around district border.