The Cost of Distance

Geography and Governance in Rural India

Sam Asher (World Bank)
Karan Nagpal (University of Oxford)
Paul Novosad (Dartmouth College)

National Institute of Public Finance and Policy

September 1, 2017

Motivation

Remoteness from urban centres matters for the rural economy

- Urban centres are large goods and labor markets
- Well-studied in literature: Fafchamps and Shilpi (2003), Donaldson and Atkin (2016), Storeygard (2016), Krishnan et al (2017)

Remoteness from urban centres matters for the rural economy

- Urban centres are large goods and labor markets
- Well-studied in literature: Donaldson and Atkin (2016), Storeygard (2016), Krishnan et al (2017)
- But also centres of administration
- Hence reduced access to public goods in more distant villages?
- This too can have negative effects on rural household outcomes

What we do

- Provide causal evidence for one of the channels through which urban distance affects rural outcomes: administration
- Use a spatial RD design that:
  - generates discontinuity in village distance to admin HQs ("administrative remoteness")
  - but not in distance to towns or highways (general market access)
- Use village-level data from India to estimate costs of admin remoteness on public goods and household outcomes
- Admin remoteness reduces public goods provision and adversely affects household outcomes
- Why? Costs, information, voice

- Administrative towns are also often important commercial or historical centres
- Hence hard to disentangle effects of administrative remoteness from other forms of remoteness
- Interesting data is often available only at district and subdistrict level - not at village level
- Challenging to think about why distance to administration affects rural outcomes - hope to learn from PMGSY, DISE

Context: District administration

- Several tiers: centre, state, district, taluk/mandal, block, Gram Panchayat
- 640 districts, 2m people on average
- District administration is powerful:
    - Colonial period: collect land revenue, enforce law and order
    - Independent India's welfare state entrusted substantial development responsibilities
    - Complicated role after 73rd and 74th amendments
    - But still an important coordinating role, implementation role for several central and state programs
- Differences in organization of district administration, in relative power of bureaucrats and politicians
    - But unifying feature: importance of district HQ town

Why distance to district HQ matters?

- From perspective of administrators:
  - Implementation costs
  - Costs of acquiring information
  - Monitoring costs: fewer inspections, official visits
- From perspective of citizens:
  - Participation in govt jobs
  - Costs of acquiring information about programs
  - Disengagement from the state
  - Costs of organizing to demand state action

Empirical Strategy Illustration
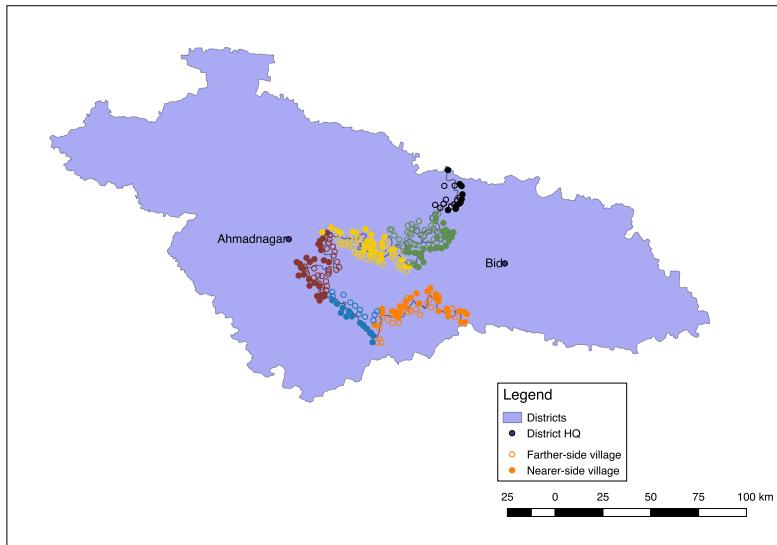
Specification

$$y_{v,d,s} = \beta_1 DistHQ_{v,d} + \beta_2 DistHighway_v + \beta_3 DistTown_v$$
$$+ f(\text{Geographic Location}_v) + \delta_d + \eta_s + \epsilon_{v,d,s}$$

- Villages within 3 km of district boundary
- Outcomes: public goods; economic activity
- Controls: distance to towns, highways, SC and ST share
- Linear controls for latitude and longitude on either side of the border segment
- Fixed effects: district (δ_d) and district border segment (η_s)
- Cluster standard errors by 50x50km grid cells to allow geographical error correlation.

Data

- Public goods
  - Population Census: roads, bus service, irrigation, schools, medical centres
- Economic outcomes
  - SECC 2012: used to derive village averages for earnings and housing quality
  - Population Census: literacy, occupation structure
- Mechanisms:
  - PMGSY program information: cost and duration for road construction
- Geocoordinates for villages, district HQs, district boundaries

Table 1: Changes in distance treatments at district boundary

	Distance to DHQ	Distance 50k town	Distance 500k town	Distance highway	Altitude(m)
Closer side of border	-31.307	0.263	0.808	0.986	-19.377
	(11.945)***	(1.730)	(1.343)	(1.015)	(12.068)
Outcome Mean	41.68	32.64	101.9	9.34	283.5
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear	Linear
N	77685	77685	77685	77685	77685

*$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Notes: This table presents regression estimates from regressing a binary variable that takes the value 1 if the village is on the more proximate side of a district border segment, and 0 otherwise, on distance to district HQ, distance to small and large towns and distance to highways. All regressions include district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Table 2: Changes in demographic treatments at district boundary

	Pop 2011	Pop change 2001-11	Migrant households	Percent SC	Percent ST
Closer side of border	-241.171	-0.013	-1.637	-0.000	0.000
	(205.721)	(0.046)	(8.037)	(0.000)	(0.000)
Outcome Mean	1170	.1653	62.29	18.68	16.26
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear	Linear
N	77685	77166	52999	77685	77685

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

Notes: This table presents regression estimates from regressing a binary variable that takes the value 1 if the village is on the more proximate side of a district border segment, and 0 otherwise, on village population in 2011, population growth between 2001 and 2011, percent of village population that is reported to belong to Scheduled Castes (SC) or Scheduled Tribes (ST) in 2011, and percentage of village households reporting at least one migrant member in the BLP Census 2002. All regressions include district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.
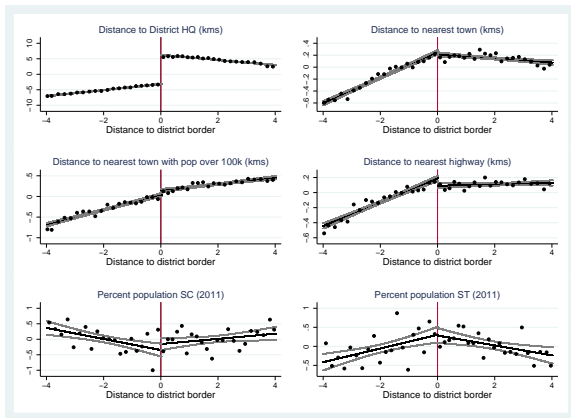
Changes at district boundary

Table 3: SECC Income Shares

	Low Income Share	Medium Income Share	High Income Share	Average Income
Distance to District HQ (kms)	0.376	0.031	-0.407	-78.487
	(0.438)	(0.344)	(0.188)**	(39.040)**
Outcome Mean	77.25	15.58	7.174	4972.784
Bandwidth	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear
N	77685	77685	77685	77685

*$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Notes: The table presents estimates from our main regression specification, where we regress household income from the SECC on distance to district HQ in kilometers, distance to nearest town, distance to highways, and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Table 4: Household assets

	Land Own Share	Solid Wall	Phone Share	Vehicle Share
Distance to District HQ (kms)	0.720	-2.223	-1.002	-0.908
	(0.657)	(0.532)***	(0.407)**	(0.344)***
Outcome Mean	55.087	51.899	66.693	18.272
Bandwidth	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear
N	77685	77685	77685	77685

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

Notes: The table presents estimates from our main regression specification, where we regress household assets on distance to district HQ in kilometers, distance to nearest town, distance to highways, and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Table 5: Occupation structure

	Nonfarm Emp	Salary Govt	Salary Private	Salary Enterprise	Income Tax
Distance to District HQ (kms)	-0.313	-0.407	-0.219	0.094	-0.376
	(0.501)	(0.094)***	(0.125)	(0.219)	(0.157)**
Outcome Mean	28.14	4.57	2.439	1.131	3.453
Bandwidth	3 km	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear	Linear
N	76891	77685	77685	77685	77685

*$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Notes: The table presents estimates from our main regression specification, where we regress household occupation structure on distance to district HQ in kilometers, distance to nearest town, distance to highways, and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Table 6: Public goods

	Paved Roads	Bus Service	Irrigation Share	Electricity
Distance to District HQ (kms)	-3.349	-1.972	-1.378	-0.501
	(1.158)***	(1.002)**	(0.469)***	(1.284)
Outcome Mean	79.86	40.47	39.04	53.66
Bandwidth	3 km	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear	Linear
N	77681	77685	74844	75706

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

Notes: The table presents estimates from our main regression specification, where we regress public goods on distance to district HQ in kilometers, distance to nearest town, distance to highways, and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Administrative remoteness affects public goods

Table 7: Schooling and literacy

	Primary School	Secondary School	Literacy Rate
Distance to District HQ (kms)	-0.407	-0.532	-0.595
	(0.783)	(0.564)	(0.25)**
Outcome Mean	84.1	12.97	56.81
Bandwidth	3 km	3 km	3 km
Fixed effect	District, Segment	District, Segment	District, Segment
Geographic polynomial	Linear	Linear	Linear
N	77668	77685	76909

*$p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Notes: The table presents estimates from our main regression specification, where we regress public goods on distance to district HQ in kilometers, distance to nearest town, distance to highways, and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Distance matters least in southern states
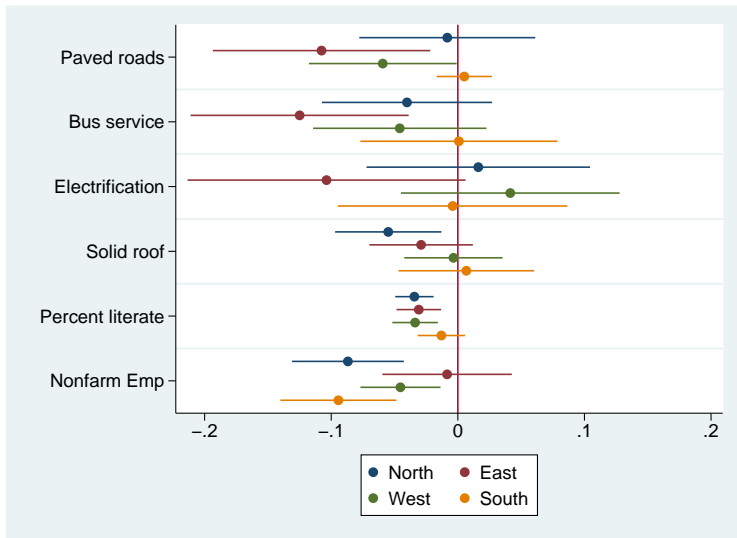
Robustness checks

- Different bandwidths around district border
- State borders
- Different geographic polynomials
- Length of border segments
- Clustering of standard errors

▸ Coefficient Plot

▸ Without state borders

▸ Quadratic   ▸ Cubic

Mechanisms

- Effects potentially driven by:
  - Higher implementation costs for public goods: vacancies, absentteeism, reduced competition among contractors: PMGSY, DISE
  - Higher monitoring costs: road quality
  - Less information: with district officials, rural residents
  - Reduced voice: less ability to organize and demand public goods from administration: GP demanded versus sanctioned funds
- One way to check: cost of PMGSY road provision
  - National program, identical standards
- Roads (slightly) more expensive to construct in villages farther from their administrative centres
  - Risk premium or reduced competition in bidding

Table 8: PMGSY outcomes

	Cost per km	Cost Overrun per km	Time per km	Time Overrun per km
Distance to District HQ (kms)	0.04	0.00	-2.00	-1.43
	(0.04)**	(0.00)	(5.09)	(4.05)
Outcome Mean	3.24	-.2005	236.8	86.47
All controls	Yes	Yes	Yes	Yes
Bandwidth	9 km	9 km	9 km	9 km
Fixed effect	Border Segment	Border Segment	Border Segment	Border Segment
Geographic Polynomial	Linear	Linear	Linear	Linear
N	49665	36982	41571	41522

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

Notes: The table presents estimates from our main regression specification, where we regress PMGSY program outcomes on distance to district HQ, distance to nearest towns, distance to highways and district and border segment fixed effects. We use linear controls for latitude and longitude. Standard errors are clustered in blocks of size 50km by 50km.

Conclusion

- Greater distance from urban centres associated with worse rural outcomes
- Distance from administrative centres reduces public goods provision
- ...and has negative impacts on rural living standards
- Effects potentially driven by higher costs of implementation, monitoring and acquiring information, and citizens' ability to organize
- But distance matters less in better-governed states

- Exploit district splits as placebo and alternate identification
- Does proximity to highways mitigate the cost penalty?
- How effects vary with age of district border
- Potential channels:
    - Voice: border effects on voter turnout, incumbency advantages
    - Information: do closer villages manipulate program information to increase odds of getting public goods?

Thank you!