The Impact of Rural Road Construction in India

Sam Asher (University of Oxford)
Paul Novosad (Dartmouth University)

June 22, 2015
Motivation

- Majority of the poor in most developing countries live in rural areas with poor linkages to trunk infrastructure
  - 33% of Indian villages had no paved road in 2001
- Insufficient research on impacts of rural roads
Question: What is the impact of village feeder roads on local economic activity and poverty?

Context: 100,000+ feeder roads built under PMGSY-India since 2000

Outcomes: Industry, employment, poverty, migration, schooling

Empirical Method: Ex-post evaluation
  - Separating selection from treatment effects
  - Regression Discontinuity approach
Launched in 2000 to connect all villages to road network

By 2011:
- 105,000 new roads, 50,000 upgrades
- 67 million people newly connected
- $320 billion disbursed

Transparent, systematic electronic record keeping: details of every habitation and road built

Successive rounds targeting unconnected habitations with populations over 1000, then 500, then 250
What does connecting a village do?
- Lowers cost of moving goods and people
- Cost of both import and export falls

Theoretical impacts are ambiguous
- Goods / Production
  - Cheaper to export (produce more in villages)
  - Cheaper to import (produce less in villages)
  - Specialize in area of comparative advantage
  - But net rural employment effects could go either way
- Migration
  - If economic opportunities in villages improve, may induce more people to stay
  - But new roads also lower cost of leaving

Places that are unconnected in 2000 have very few economic opportunities
Few reliable studies on road impacts (Khandker et al 2009, Mu and van de Walle (2009), Casaburi et al (2014))

Challenges:
- Selection vs. Treatment
- Sample Size
Roads are not built at random
- Places where roads are built may be:
  - Politically Connected (positive bias)
  - Have high value economic opportunities (positive bias)
  - Be poorer than other places (negative bias)

Comparison of newly connected and unconnected places measures sum of treatment and selection
- All road studies find significant pre-treatment selection
- Controlling for observables may not be sufficient
Roads are not built at random
- Places where roads are built may be:
  - Politically Connected (positive bias)
  - Have high value economic opportunities (positive bias)
  - Be poorer than other places (negative bias)

Comparison of newly connected and unconnected places measures sum of treatment and selection
- All road studies find significant pre-treatment selection
- Controlling for observables may not be sufficient

Village with ineffective governance
Village with effective governance
Road programs are usually small
- Casaburi 2014, Sierra Leone, n=31
- Mu and van der Walle, Vietnam, n=95
- Exception: Khandker 2009, Bangladesh, n=1500

Limited ability to do inference

Selection issues may be greater when many are excluded
Our Approach

- PMGSY: 100,000+ new roads from 2001-present
- Census microdata at national scale
  - Administrative data from government road database
  - Employment and category of all non-farm establishments
  - Village-level demographics
  - Education, enrolment and occupation
  - Poverty indicators: Assets, household income, land use
  - Unavailable: Wages, agricultural outcomes, prices

- Natural Experiment: Regression Discontinuity
Natural Experiment: Regression Discontinuity

- Gold Standard: Randomized Control Trial
  - If roads are allocated at random, selection effect goes away
  - Average outcome difference between road and no road is unbiased estimate of treatment effect

- Natural Experiment
  - First wave of PMGSY roads targeted villages 1000 people or more
  - Second wave: 500 or more

- Villages with population just over population threshold more likely to be treated
  - Compare outcomes in villages with population 999 to 1001
  - Selection mechanism unlikely to be correlated with other confounders
Advantages of Administrative / Census Data

- High external validity
- Administrative data may be more reliable
- Unlimited scope for joining across types of data
  - Example: 1% sample on agriculture, 1% sample on employment
- Wide scope for analysis of treatment heterogeneity
  - Who benefits most / least?
  - Every affected subgroup is represented
- New methods in CS / Statistics can significantly reduce sample noise
Dependent variable: Share of households with any durables by 2011

<table>
<thead>
<tr>
<th>Model</th>
<th>New road before 2011</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-0.027</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subdistrict Fixed Effects</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>23663</td>
</tr>
<tr>
<td>r2</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* * p < 0.10, ** p < 0.05, *** p < 0.01
Dependent variable: Share of households with any durables by 2011

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New road before 2011</td>
<td>-0.027</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>(0.003)***</td>
<td>(0.004)***</td>
</tr>
<tr>
<td>Sample</td>
<td>Full</td>
<td>Baseline No Road</td>
</tr>
<tr>
<td>Subdistrict Fixed Effects</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>23663</td>
<td>16683</td>
</tr>
<tr>
<td>r2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01
Dependent variable: Share of households with any durables by 2011

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New road before 2011</td>
<td>-0.027</td>
<td>-0.026</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.003)***</td>
<td>(0.004)***</td>
<td>(0.003)**</td>
</tr>
<tr>
<td>Assets at baseline</td>
<td></td>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.000)***</td>
</tr>
<tr>
<td>Sample Subdistrict Fixed Effects</td>
<td>Full No</td>
<td>Baseline No Road No</td>
<td>No Road No</td>
</tr>
<tr>
<td>N</td>
<td>23663</td>
<td>16683</td>
<td>16683</td>
</tr>
<tr>
<td>r2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01
Dependent variable: Share of households with any durables by 2011

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New road before 2011</td>
<td>-0.027</td>
<td>-0.026</td>
<td>-0.007</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.003)***</td>
<td>(0.004)***</td>
<td>(0.003)**</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Assets at baseline</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.000)***</td>
<td>(0.000)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subdistrict Fixed Effects</td>
<td>Full No</td>
<td>Baseline No Road No</td>
<td>No Road No</td>
<td>No Road Yes</td>
</tr>
<tr>
<td>N</td>
<td>23663</td>
<td>16683</td>
<td>16683</td>
<td>16683</td>
</tr>
<tr>
<td>r2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*p < 0.10, **p < 0.05, ***p < 0.01
Regression Discontinuity Estimate

- Share of villages connected by 2011 vs. Population 2001
- Share of households with any assets in list vs. Population 2001
Interpretation: Migration

Sex Ratio Age 16-35

Population 2001
Conclusion

- Preliminary results
  - Roads allow people to work outside village
  - Including some who would have been in school
  - Migrants appear to be richer than non-migrants
  - Effects vary by state

- RD estimates significantly different from conventional OLS estimates
  - Suggests significant selection into treatment
  - Naive evaluation will differ biased estimates
Conclusion

- RD estimates are unbiased
  - Require huge amount of data
  - Less transparent for policy-makers
- Migration major confounder in infrastructure projects
  - Revealed preference suggests people better off with option to migrate
  - But most benefits may accrue to people leaving
  - So static surveys may miss winners