Did railways affect literacy? Evidence from India

Latika Chaudhary and James Fenske

Naval Postgraduate School, University of Warwick

Railroads and Schooling in India

Research Question

Did the extension of colonial railroads lead to a demand-driven increase in Indian education?

- By 1900, the rail network in colonial India was the fourth largest in the world, covering more than 40,000 kilometers across more than 200 districts (Bogart and Chaudhary, 2016).
- In contrast, education was an insignificant line item in the government budget (1.7%) and in 1891 only 9.6% of primary school-age children were in school (Chaudhary, 2016).
- Two broad views on why India had poor education outcomes.
  - Demand for basic education was low in a rural agricultural economy like India (e.g., Great Britain, 1929).
  - Insufficient public funding of education (e.g., Chaudhary, 2016).

What do we do?

- Using decennial census data on Indian literacy from 1881 to 1921, we estimate the effect of railroads on total, male, female, and English literacy at the district level.
- First, synthetic panel, approach: estimate the effect of exposure to railroads using variation across cohorts within districts using the censuses of 1911 and 1921, including district, cohort $\times$ province, and census year $\times$ province fixed effects.
- Second, cross-sectional approach: we construct one instrument using a 1852 plan based on low-cost routes and another using a spanning tree connecting 54 military cantonments c. 1864.

Results

- Synthetic panel: A standard deviation increase in railroad exposure (17 years) increases total literacy by 0.29 standard deviations for total, 0.31 for male, and 0.25 for male English literacy. Cross-sectional results are similar, with additional effects on female literacy.
- Proximate mechanism: Using novel data on primary and secondary enrolment at the district level, we show railways increased secondary enrolment.
- Deeper mechanisms: no evidence that agricultural income is a mediator; rather income taxes, urbanisation, and service sector employment are key mediators. These proxy for non-agricultural income and the returns to skill.

Contribution

- On the effects of transportation infrastructure (e.g. Adukia et al., 2020; Aggarwal, 2018):
  - We find that the effects of colonial railroads are driven more by increasing returns to education and higher non-agricultural income rather than by income or substitution effects of rising agricultural income.
  - We show that the impacts of transportation infrastructure on human capital have not been limited to modern economies.
- On the effects of demand and supply in explaining schooling (e.g. Glewwe and Muralidharan, 2016)
  - Our paper shows that one of the biggest infrastructure expansions, railroads, had positive effects on literacy and enrolment in India. Yet, these effects are modest and hence not cost-effective if we consider them against increased public funding of education.
- On Indian railroads (e.g. Hurd, 1983; Bogart and Chaudhary, 2016; Donaldson, 2018):
  - There were positive effects on schooling, though these favoured men, English literacy, and secondary enrolment.

1. Introduction
2. Historical Background
3. Data
4. Estimation Strategy
5. Results
6. Mechanisms
7. Conclusion

1 Introduction

2 Historical Background

3 Data

4 Estimation Strategy

5 Results

6 Mechanisms

7 Conclusion

Literacy

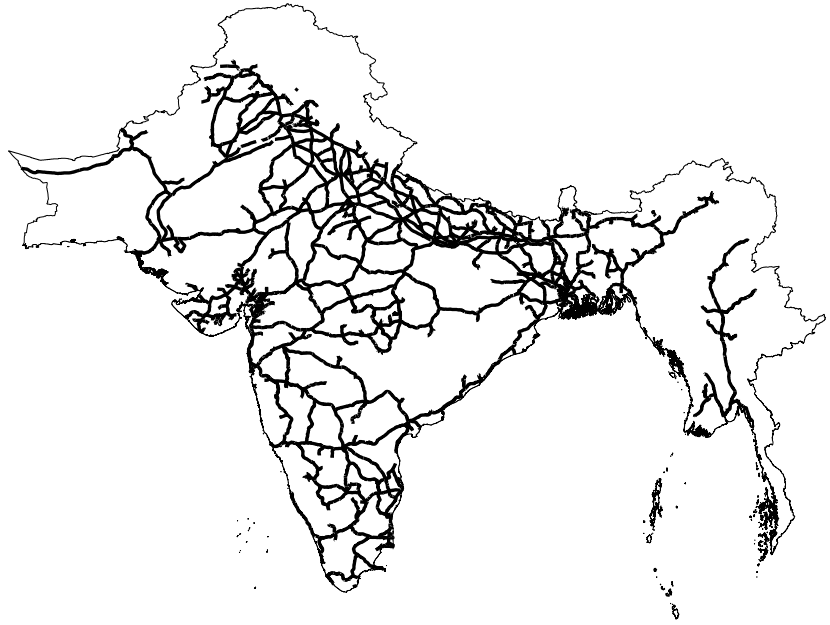
- The raw data suggest literacy was low but increased from 1881 to 1921. Male literacy increased from 6% in 1881 to 12% in 1921, while female literacy increased from under 1% to 2% over the same decades (Census of India, 1931).
- Enrolment rates increased faster than literacy, from one in ten children attending school in 1891 to just over 1 in 5 in 1921 (Chaudhary, 2016).
- Decentralisation of the public funding of education in the 1870s and 1880s led to stark differences in public spending across provinces.
- Yet, these differences in public spending across provinces did not translate into differences in school outputs, namely enrolment or literacy.
- Apart from regional differences, there were large differences in schooling by caste and religion – Brahmans, Christians, and Parsis, for example, had higher literacy rates.

Railroads

- The first passenger line opened in 1853, connecting Bombay to Thane.
- The important ports were connected to the interior before 1881.
- Many lines crossed the densely populated Indo-Gangetic plain with fewer interior lines in the Deccan plateau.
  - Early proposals such as the Kennedy plan in 1852 called for lines parallel to the coast in order to economise on costs.
- Although British firms built the railways, the GOI dictated route placement.
- What guided their decisions? Military, commercial, and famine concerns were cited as the main drivers in official correspondence (Hurd, 1983).
- So: we could expect either positive or negative selection. This motivates the synthetic panel and IV approaches.

Figure: Rail Network 1881-1921

1881



1921

1 Introduction

2 Historical Background

3 Data

4 Estimation Strategy

5 Results

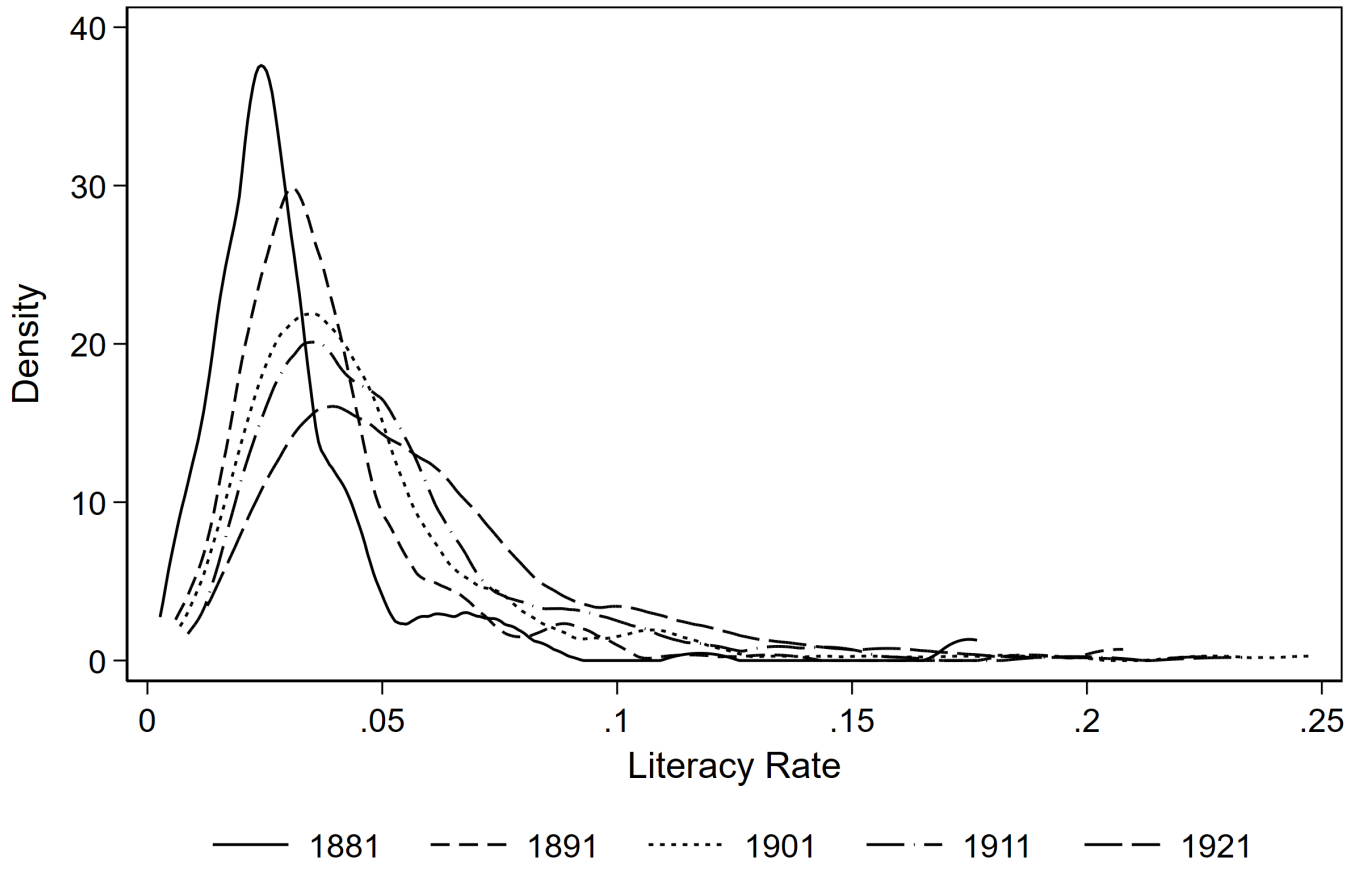
6 Mechanisms

7 Conclusion

Literacy

- We use the colonial censuses of 1881-21 to measure literacy.
- Because measures differ across censuses before 1911, we focus on inter-cohort differences in the 1911 and 1921 censures.
- We use district gazetteers to produce new data on primary and secondary enrolment between 1894 and 1911.
  - Many secondary schools had attached primary classes, so some primary aged children will, then, be included in secondary enrolment.
  - Such primary classes were of higher quality than regular vernacular primary schools.

Figure: Distribution of Total Literacy



Distribution truncated at 25%. Fewer than 1% of observations are above this cutoff.

Railroads (1/2)

- We define two alternative measures of railroad years assuming railroads affect literacy only up to the beginning or end of elementary school.
- Our first measure is the number of years a railroad has been operating in a district minus the number of years since the youngest member of a cohort would have regularly begun elementary school, i.e. at age 6.
- Denote the number of years since schooling began as $y(c)$. For cohorts aged 20 and above, $y(c)$ is 14. For cohorts aged 15-20, it is 9. For cohorts aged 10-15, it is 4. For cohorts aged below 10, it is 0.
- For cohort c, $y(c)$ years since schooling began, in district d, with a railroad that opened in year r, measured in census year t, our first is:

$$RailroadYears_{cdt} = \begin{cases} \max\{t - r - y(c), 0\} & \text{if } r \leq t, \\ 0 & \text{if } r > t. \end{cases}$$

Railroads (2/2)

$$RailroadYears_{cdt} = \begin{cases} \max\{t - r - y(c), 0\} & \text{if } r \leq t, \\ 0 & \text{if } r > t. \end{cases}$$

- Suppose a railroad opens in a district in 1901. In the 1911 census, the railroad has been active for 10 years.
- In the cohort aged 10-15, the youngest member of that cohort was 6 in 1907 and has not yet turned 12.
- By our first measure, the railroad exposure for that cohort is $\max\{1911 - 1901 - 4, 0\} = 6$ years.
- If, instead, the railroad had been built in 1912, exposure for all cohorts would be zero in the 1911 census.
- In the cross section, we simply measure how long a railway has been present.

1 Introduction

2 Historical Background

3 Data

4 Estimation Strategy

5 Results

6 Mechanisms

7 Conclusion

Synthetic Panel

$$\ln(\mathit{LiteracyRate}_{cdt}) = \beta \mathit{RailroadYears}_{cdt} + \theta_d + \delta_p \times \eta_t + \delta_p \times \gamma_c + \epsilon_{cdt}$$

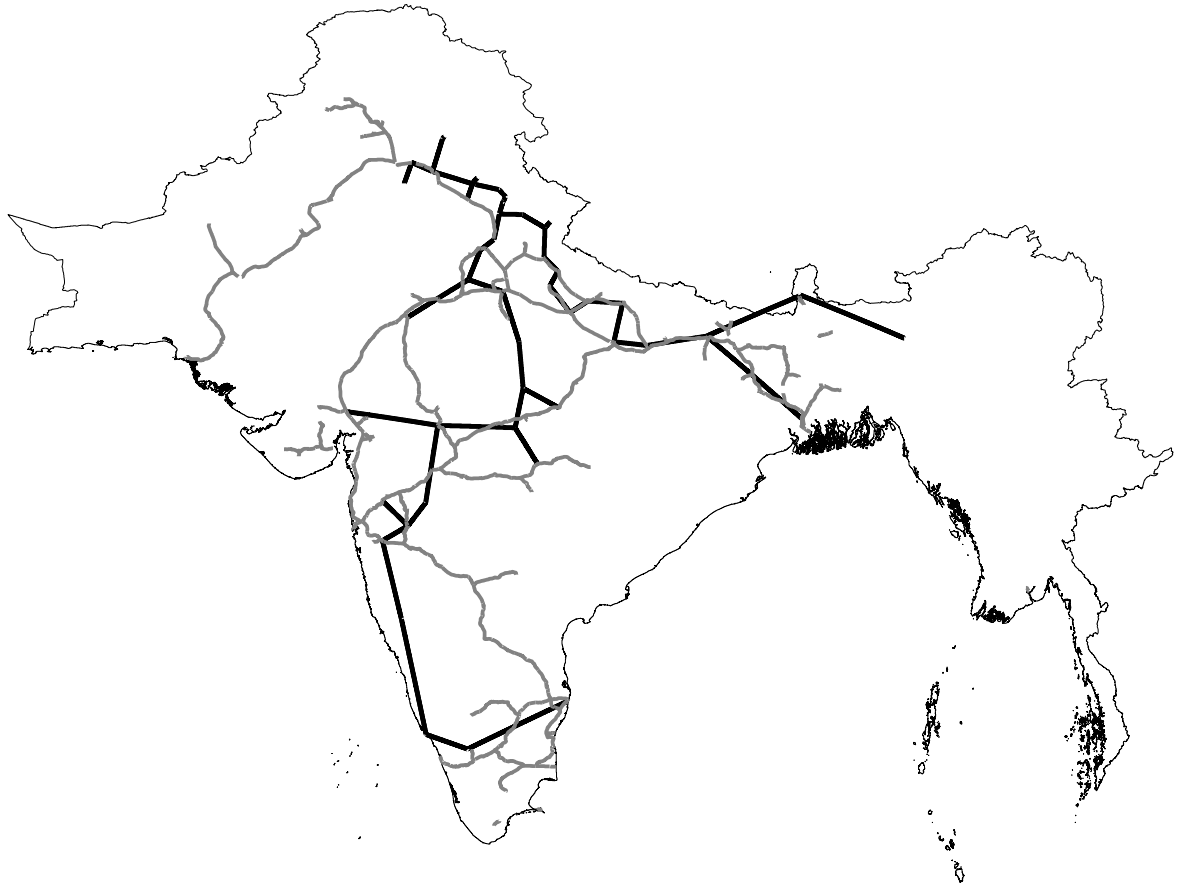
- $\mathit{LiteracyRate}_{cdt}$ is literacy for cohort c in district d and census year t.
- $\mathit{RailroadYears}_{cdt}$ measures the cumulative years of railroad exposure for cohort c in district d in year t.
- We estimate the model for $t \in \{1911, 1921\}$ and cohort $c \in \{0-10, 10-15, 15-20, 20+\}$.
- We control for district fixed effects, θ_d, province $\times$ year fixed effects $\delta_p \times \eta_t$, and province $\times$ cohort fixed effects $\delta_p \times \gamma_c$.
- We cluster standard errors by district to account for serial correlation over time.
- As a robustness check, we estimate Conley (1999) standard errors that account for spatial correlation with cutoffs ranging from 200km to 500km.

Cross-Section

$$\ln(LiteracyRate_{dt}) = \beta RailroadYears_{dt} + \gamma' x_{dt} + \delta_p + \epsilon_{dt}$$

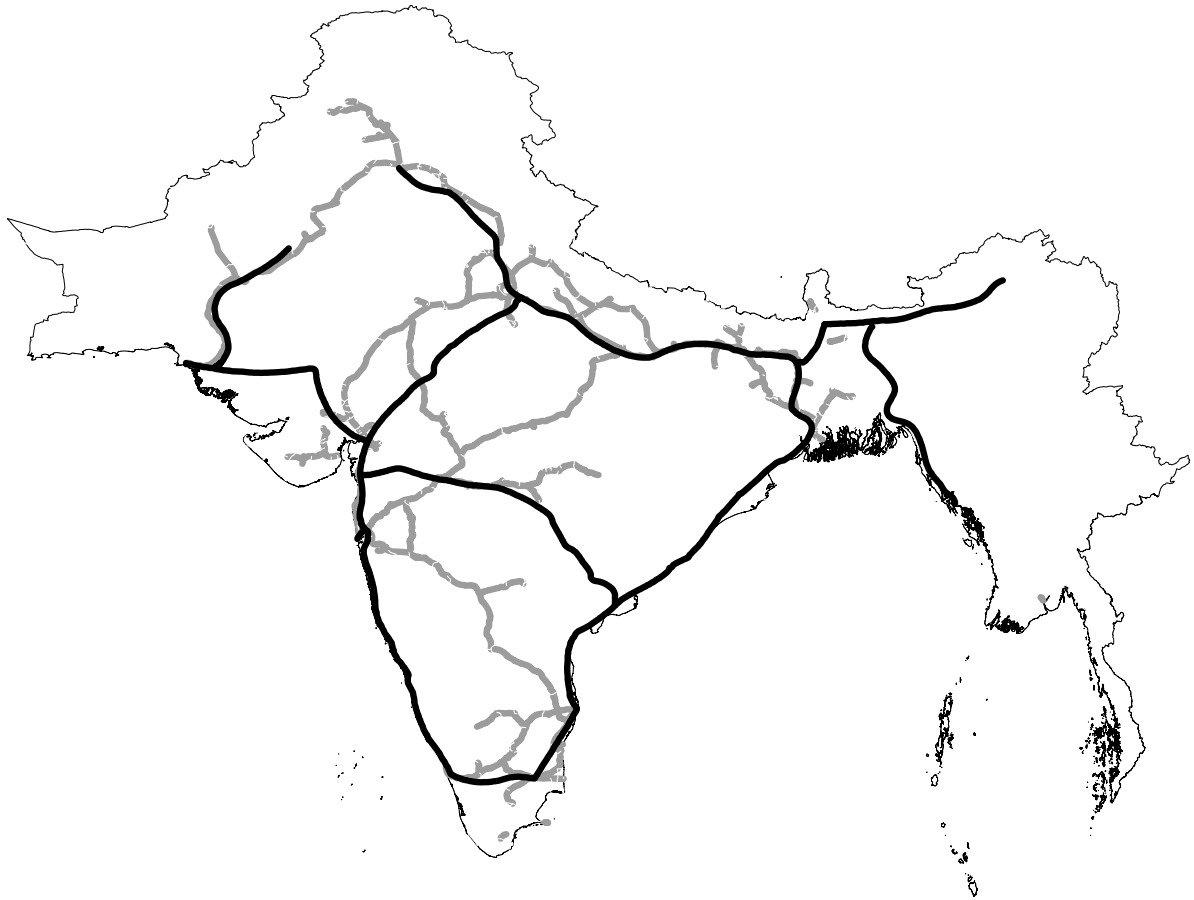
- We estimate this regression separately for $t \in \{1881, 1891, 1901, 1911, 1921\}$.
- $\ln(LiteracyRate_{dt})$ is the log literacy rate in district d in year t.
- $RailroadYears_{dt}$ is the number of years district d in year t has had a railroad.
- x_{dt} includes the GIS controls, pre-rail urbanisation and social controls.
- We also include province fixed effects captured by δ_p.
- We use robust standard errors in the baseline and Conley (1999) standard errors for robustness.
- Instrument 1: Log of (one plus) distance from a minimum tree spanning 54 military cantonments that existed in 1864.
- Instrument 2: Log of (one plus) distance from Major Kennedy's 1852 plan, which favored low-cost routes.

Figure: Map of Military Cantonment Spanning Tree



Spanning tree drawn in black. 1881 railway network drawn in grey.

Figure: Map of 1852 Kennedy Plan



1852 Kennedy Plan drawn in black. 1881 railway network drawn in grey.

1 Introduction

2 Historical Background

3 Data

4 Estimation Strategy

5 Results

6 Mechanisms

7 Conclusion

Synthetic Panel Results

- Standardised coefficients range from 0.29 standard deviations for total and 0.31 standard deviations for male literacy.
- We find smaller effects on male English compared to male non-English literacy with standardised coefficients at 0.25 for English and 0.34 for non-English literacy.
- Unlike males, we find small and insignificant effects of railroads on female literacy, female English and female non-English literacy.
- Appendix robustness: discard the (most mis-measured) cohort aged 20 and above; use exposure up to age 12; use the 1901 census.

Synthetic Panel: Cohort, District and Year Fixed Effects

	(1)	(2)	(3)
	Total	Male	Female
	Literacy		
Cohort Years	0.0202***	0.0224***	0.0079
of Railroad Exposure	(0.0070)	(0.0071)	(0.0078)
Obs.	1,609	1,609	1,608
	English Literacy		
Cohort Years	0.0234***	0.0266***	0.0050
of Railroad Exposure	(0.0078)	(0.0086)	(0.0079)
Obs.	1,598	1,597	1,536
	Non-English Literacy		
Cohort Years	0.0212***	0.0235***	0.0080
of Railroad Exposure	(0.0074)	(0.0075)	(0.0081)
Obs.	1,607	1,607	1,606
Years	1911-1921	1911-1921	1911-1921

Cross-Sectional Results

- Including controls and province fixed effects, standardised coefficients range from 0.1 to 0.22 standard deviations, with those for English and female literacy being on the higher end of the range.
- IV estimates: For 1901, standardised coefficients are 0.54 for English literacy, 0.43 for female literacy, and 0.36 for male literacy.
- Appendix robustness: Conley standard errors; grid cell fixed effects; nearest neighbor matching.
- Why do we now find effects for women? This may be due to differences in the LATEs estimated by the two methods, exposure at later ages, and measurement error with fixed effects.

These effects are small

- First: benchmark our results against those in Atack et al. (2012).
  - They estimate the effect of railroads on individual school enrolment in the United States.
  - Increasing rail access across US counties in the 1850s predicts 56% of the increase in mean school enrolment between 1850 and 1860.
  - In our case, increasing exposure to railroads between 1881 and 1891 predicts 16% of the actual increase in literacy.
- Second: think about cost effectiveness.
  - Chaudhary (2010b) finds it would have cost the colonial government roughly 3 rupees to make an additional person literate using causal estimates of public education spending on literacy.
  - Using data on construction costs from Bogart and Chaudhary (2016), we estimate a cost of 606 rupees to make one additional person literate.

1 Introduction

2 Historical Background

3 Data

4 Estimation Strategy

5 Results

6 Mechanisms

7 Conclusion

Mediators

- Enrolment is a proximate mechanism.
  - We find positive effects on secondary enrolment, which in our data sources includes higher quality primary classes and English instruction.
- Agricultural income
  - Agricultural income does not predict schooling; low-skilled agricultural wages do not respond to railroads.
- Non-Agricultural Income, Urbanisation and Returns to Education
  - Income taxes mediate 30% to 46% of the effects of railroads on literacy, and 25% to 43% of the effects on secondary enrolment.
  - Urbanisation mediates between 38% and 48% of the effects on total literacy, and a smaller share of secondary enrolment at 9% to 16%.
  - Service sector employment also appears to partially mediate the results, but less so than urbanisation and income taxes. It mediates anywhere from 6% to 16% of the effect of railroads on literacy and secondary enrolment.
- In the appendix, we rule out mediating roles for migration, Europeans, missions, railroad workers, and recruitment of soldiers.

Conclusion

- We study the effects of railroads on Indian literacy and enrolment using district-level data from 1881 to 1921.
- We find positive and significant effects of railroads on male and English literacy.
- Our results are robust in both panel models where we exploit variation in railroad exposure across cohorts within districts and in cross-sectional models where we control for the endogeneity of railroad exposure using instrumental variables. Railroads lead to greater literacy via higher secondary enrolment.
- We find no evidence that agriculture is an important mediator.
- Rather, non-agricultural income, urbanisation and service sector employment are key mediators of the link between railroads and higher schooling.
- Railroads generated positive spillovers on education, but their effects were concentrated and not broadly shared.