Railroads of the Raj:

Estimating the Impact of Transportation Infrastructure

Dave Donaldson

London School of Economics

Transportation Infrastructure

- Empirical Questions:
  1. How large are the economic benefits of transportation infrastructure projects (which aim to reduce trade costs)?
  2. What economic mechanisms explain these benefits?
- Motivation:
  - 20 percent of 2007 World Bank loans allocated to transportation infrastructure projects
  - Widespread policy initiatives aim to reduce trade costs more generally: tariffs, corruption, red tape

Approach of This Paper

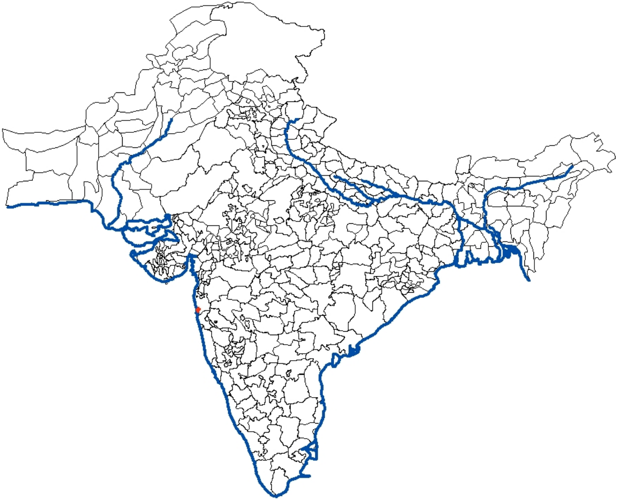
- Study large improvement in transportation technology—Railroads—in setting with best possible data—colonial India ("the Raj")

Approach of This Paper

- Study large improvement in transportation technology—Railroads—in setting with best possible data—colonial India ("the Raj")
- Construct new dataset on Indian economy before and after the railroads
    - Output, prices, internal and external trade
    - District-level ($N = 239$), annual 1861-1930

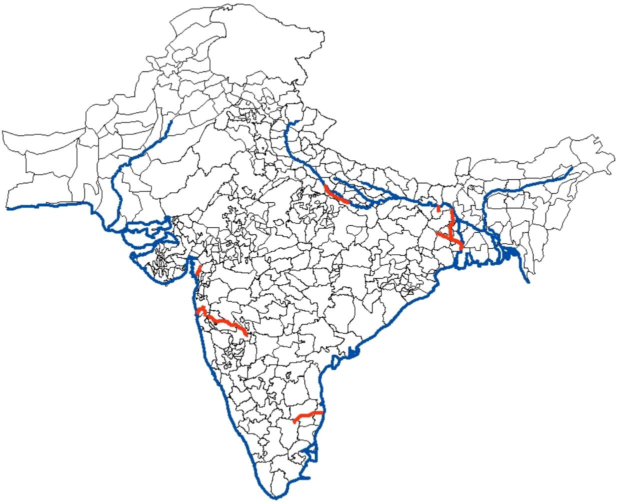
Indian Transportation Network: 1853

Eve of railroad age: first track in 1853



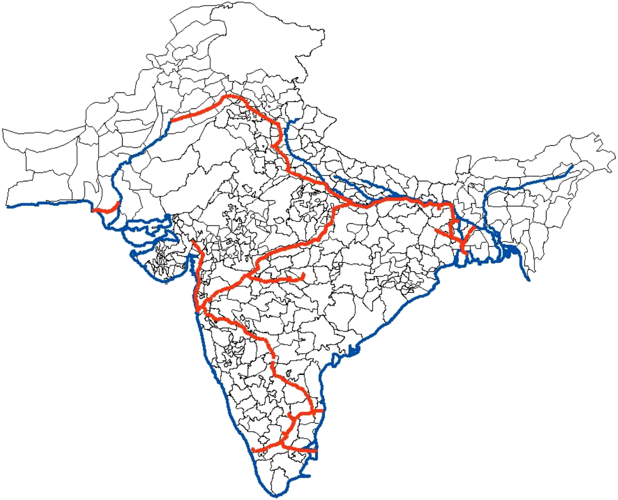
Indian Transportation Network: 1860

Each railroad 'pixel' coded with its year of opening

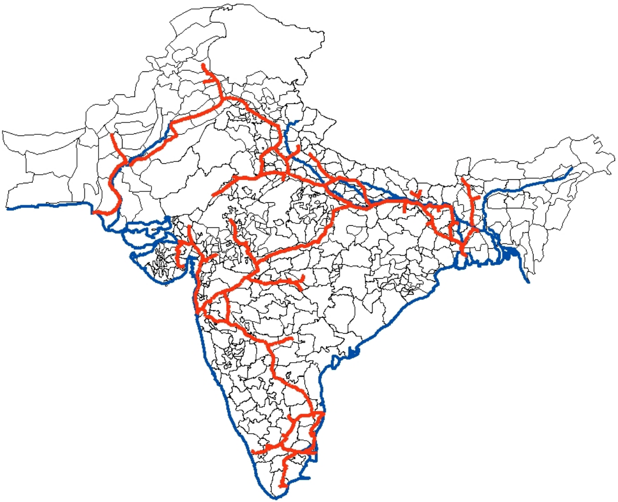


Indian Transportation Network: 1870

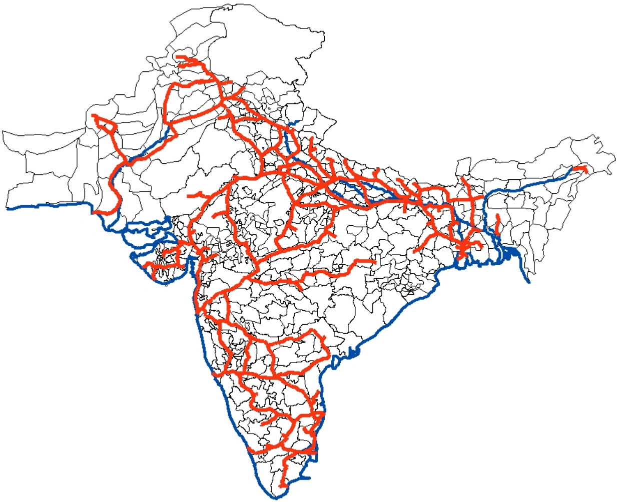
Seven provincial capitals connected



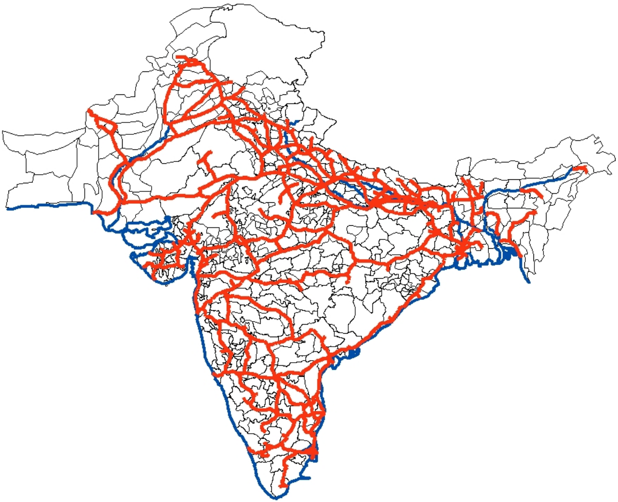
Indian Transportation Network: 1880



Indian Transportation Network: 1890

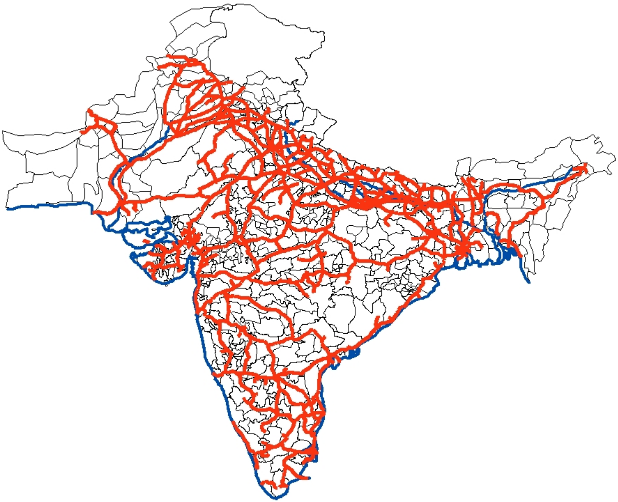


Indian Transportation Network: 1900

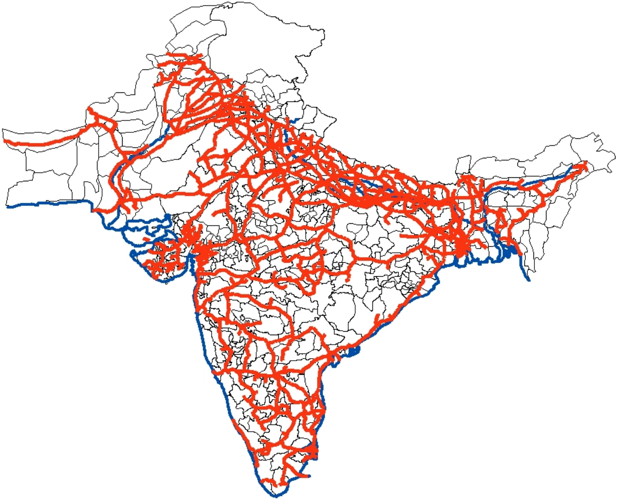


Indian Transportation Network: 1910

4th largest railroad network in the world

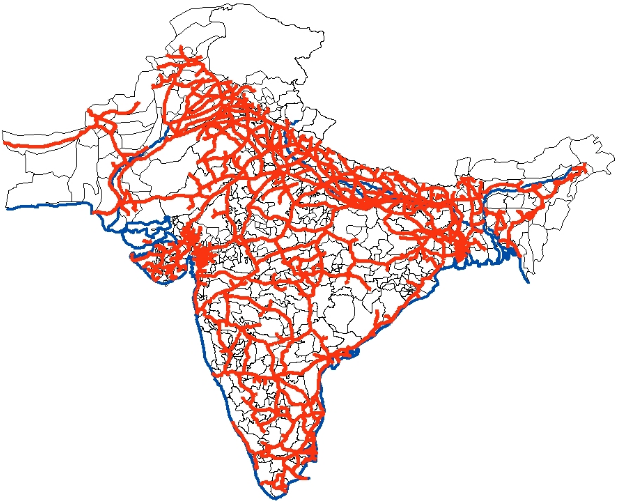


Indian Transportation Network: 1920



Indian Transportation Network: 1930

Network in 2009 is effectively that in 1930. 67,247 km of line open.



Approach of This Paper

- Study large improvement in transportation technology—Railroads—in setting with best possible data—colonial India ("the Raj")
- Construct new dataset on Indian economy before and after the railroads
  - Output, prices, internal and external trade
  - District-level ($N = 239$), annual 1861-1930

Approach of This Paper

- Study large improvement in transportation technology—Railroads—in setting with best possible data—colonial India ("the Raj")
- Construct new dataset on Indian economy before and after the railroads
  - Output, prices, internal and external trade
  - District-level ($N = 239$), annual 1861-1930
- Use GE trade model (based on Eaton and Kortum, 2002) to guide empirical approach
  - Comparative advantage (Ricardian) model of trade
  - Trade costs are primitive in model
  - Model makes 6 testable predictions

Step	Did railroads...	Result
1		
2		
3		
4		
5		
6		

Step	Did railroads...	Result
1		
2		
3		
4		
5		
6		

Step	Did railroads...	Result
1	...reduce trade costs (and price gaps)?	Yes
2		
3		
4		
5		
6		

Step	Did railroads...	Result
1	...reduce trade costs (and price gaps)?	Yes
2	...expand trade?	Yes
3		
4		
5		
6		

Step	Did railroads...	Result
1	...reduce trade costs (and price gaps)?	Yes
2	...expand trade?	Yes
3		
4	...raise real income level?	Yes
5		
6		

Step	Did railroads...	Result
1	...reduce trade costs (and price gaps)?	Yes
2	...expand trade?	Yes
3		
4	...raise real income level?	Yes
5		
6	...promote (static) gains from trade?	

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	
2	...expand trade?	Yes	
3			
4	...raise real income level?	Yes	
5			
6	...promote (static) gains from trade?		

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	
3			
4	...raise real income level?	Yes	
5			
6	...promote (static) gains from trade?		

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3			
4	...raise real income level?	Yes	
5			
6	...promote (static) gains from trade?		

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3			
4	...raise real income level?	Yes	
5			
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	... Increase price responsiveness?	Yes	
4	...raise real income level?	Yes	
5			
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	... Increase price responsiveness?	Yes	
4	...raise real income level?	Yes	
5	... reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Outline of Talk

Historical Background

Model: 4 Predictions

4 Empirical Steps

- Step 1: Railroads and Trade Costs
- Step 2: Railroads and Trade Flows
- Step 4: Railroads and Real Income
- Step 6: Railroads and Gains from Trade

Conclusion

Outline of Talk

The Colonial Indian Economy

- Primarily agricultural:
  - 66 % of GDP in 1900 (Heston 1983)
  - Factory-based manufacturing extremely small: 1-3 % of GDP
- Agriculture was primarily rain-fed: 14 % irrigation in 1900
- $\Rightarrow$ Focus on agriculture, and use rainfall as exogenous (and observable) shock to productivity

Transportation in Colonial India

- Pre-rail transportation (Deloche 1994, 1995):
  - Roads: bullocks, 10-30 km per day (ie 2-3 months to port)
  - Rivers: seasonal, slow
  - Coasts: limited port access for steamships
- Railroad transportation:
  - Faster: 600 km per day
  - Safer: predictable, year-round, limited damage, limited piracy
  - Cheaper:
    - $\sim 4.5\times$ cheaper than roads
    - $\sim 3\times$ cheaper than rivers
    - $\sim 2\times$ cheaper than coast

Outline of Talk

Historical Background

Model: 4 Predictions

4 Empirical Steps

- Step 1: Railroads and Trade Costs
- Step 2: Railroads and Trade Flows
- Step 4: Railroads and Real Income
- Step 6: Railroads and Gains from Trade

Conclusion

Model Set-up

- Multi-sector version of Eaton and Kortum (2002)—general equilibrium with:
  - Many (≥ 2) regions
  - Many (≥ 2) goods
  - Trade costs $T \in [1, \infty)$

- K goods (e.g. rice, wheat):
  - indexed by k
  - each available in continuum of varieties (j)

- D regions (districts, foreign countries)
  - o = origin
  - d = destination

- Static model

Model Environment

- Technology: $q_o^k(j) = L_o^k \; z_o^k(j) \qquad p_{oo}^k(j) = \frac{r_o}{z_o^k(j)}$

$$z_o^k(j) \sim F_o^k(z) = \exp(-A_o^k \, z^{-\theta_k})$$

- Tastes: $\ln U_o = \sum_{k=1}^K \left(\frac{\mu_k}{\varepsilon_k}\right) \ln \int_0^1 (C_d^k(j))^{\varepsilon_k} dj$

- Trading: iceberg trade costs $T_{od}^k \geq 1, \; T_{oo}^k = 1$

$$\Rightarrow p_{od}^k(j) = T_{od}^k \, p_{oo}^k(j)$$

▸ Prices ▸ Adding Time

Prediction 1: Trade Costs

- Prediction 1: If good 'o' can only be made in one region (region o) but this good is consumed elsewhere (region d), then:

$$\ln p_d^o - \ln p_o^o = \ln T_{od}^o$$

- Useful: allows estimation of how railroads affect (unobserved) trade costs T_{od}^o

Prediction 2: Trade Flows

- Prediction 2: Exports take gravity form:

$$\pi_{od}^{k} \equiv \frac{X_{od}^{k}}{X_{d}^{k}} = \lambda^{k} A_{o}^{k} \left(r_{o} T_{od}^{k}\right)^{-\theta_{k}} \left(p_{d}^{k}\right)^{\theta_{k}}$$

- Useful: allows estimation of
  - unknown parameters θ_k
  - unknown relationship between (unobserved) A_o^k and rainfall shocks: $\ln A_o^k = \kappa RAIN_o^k$

Prediction 4: Real Income Levels

- Welfare (of representative agent owning unit of land) is equal to real income:

$$V(\mathbf{p}_o, r_o) = \frac{r_o}{\widetilde{P}_o} = \frac{Y_o}{L_o \widetilde{P}_o}$$

- Prediction 4: Real income ($\frac{Y}{L\widetilde{P}}$) and trade costs (T) around a symmetric equilibrium:

$$\frac{d\left(\frac{Y_o}{L_o \widetilde{P}_o}\right)}{dT_{od}^k} < 0$$

Prediction 6: Sufficient Statistic Property

- Prediction 6: Despite complex GE interactions, real income can be written as:

$$\ln\left(\frac{Y_o}{L_o \widetilde{P}_o}\right) = \Omega + \sum_k \frac{\mu_k}{\theta_k} \ln A_o^k - \sum_k \frac{\mu_k}{\theta_k} \ln \pi_{oo}^k$$

- Useful: 'Autarkiness' (π_{oo}^k) is a sufficient statistic for all of the effects of the railroad network on real income

▸ Prediction 3 ▸ Prediction 4 (b) ▸ Prediction 5

Outline of Talk

Historical Background

Model: 4 Predictions

4 Empirical Steps

- Step 1: Railroads and Trade Costs
- Step 2: Railroads and Trade Flows
- Step 4: Railroads and Real Income
- Step 6: Railroads and Gains from Trade

Conclusion

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	...reduce price responsiveness?	Yes: to ≈ 0	Model evaluation
4	...raise real income level?	Yes	
5	...reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Conditions Required for Prediction 1

Prediction 1: $\ln p^o_{dt} - \ln p^o_{ot} = \ln T^o_{odt}$

- Good differentiated by source

- Good consumed widely at regions away from source

- Free spatial arbitrage

- Homogeneous good (Broda and Weinstein, 2008)

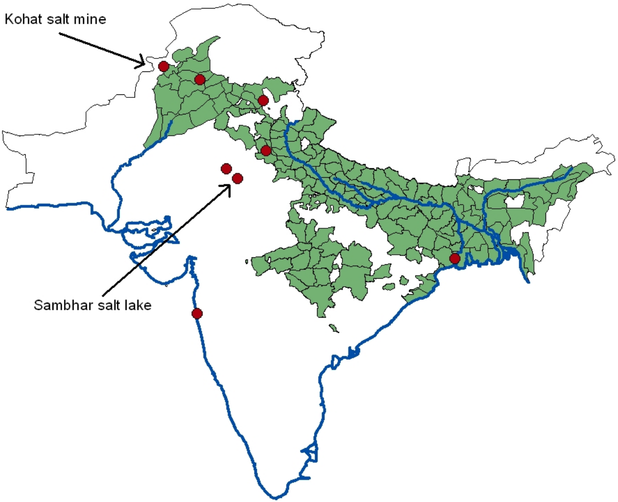
Conditions (Plausibly) Satisfied by Salt

Prediction 1: $\ln p^{o}_{dt} - \ln p^{o}_{ot} = \ln T^{o}_{odt}$

- Good differentiated by source
  - Each type could only be made in one location
  - "Kohat salt" vs. "Sambhar salt" (and 6 others)
- Good consumed widely at regions away from source
  - Biologically essential
- Free spatial arbitrage
  - Sold to unrestricted trading sector at 'factory' gate
- Homogeneous good (Broda and Weinstein, 2008)

8 Salt Sources and 125 Sample Districts

Annual data, 1861-1930



Empirical Specification

- Theory: $\ln p^o_{dt} = \ln p^o_{ot} + \ln T^o_{odt}$
- Empirical version:

$$\ln p^o_{dt} = \overbrace{\beta^o_{ot}}^{=\ln p^o_{ot}} + \overbrace{\beta^o_{od} + \phi^o_{od} t + \delta \ln LCR(\mathbf{R}_t; \boldsymbol{\alpha})_{odt} + \varepsilon^o_{dt}}^{=\ln T^o_{odt}}$$

- $LCR(\mathbf{R}_t, \boldsymbol{\alpha})_{odt}$: 'lowest-cost route'

Lowest-cost Route: $LCR(\mathbf{R}_t; \boldsymbol{\alpha})_{odt}$

- Two inputs:
  1. Model full transport system (rail, road, river, coast) in each year as a network: $\mathbf{R}_t$
     - 7651 nodes
     - $\sim$ 3 million links out of potential $\sim$ 59 million links (7651$\times$7651)
     - ▸ Network
  2. Per-unit distance trade cost of each mode: $\boldsymbol{\alpha}$
     - $\boldsymbol{\alpha} \doteq (\alpha^{rail} = 1, \alpha^{road}, \alpha^{river}, \alpha^{coast})$
- Assume: Perfectly competitive trading sector, no fixed costs of trading, no congestion, traders know $(\mathbf{R}_t, \boldsymbol{\alpha})$, traders choose cheapest route

Lowest-cost Route: $LCR(\mathbf{R}_t; \boldsymbol{\alpha})_{odt}$

- Conditional on $\boldsymbol{\alpha}$, solve for lowest-cost route over $\mathbf{R}_t$ for each *o*-*d* pair (in each year t):
  - Computationally feasible, due to Dijkstra's 'shortest path' algorithm
- Search over $(\delta, \boldsymbol{\alpha})$ to minimize squared residuals of price equation $\Rightarrow (\widehat{\delta}, \widehat{\boldsymbol{\alpha}})$

Trade Costs: Baseline Results

$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \delta \ln LCR(\mathbf{R}_t; \boldsymbol{\alpha})_{odt} + \varepsilon^o_{dt}$

Dependent variable: log destination salt price	OLS (1)
Log distance to source along lowest-cost route (ie LCR($\mathbf{R_t}$, α))	0.135 (0.038)***
Mode-wise relative marginal costs	
Rail: (ie α^{rail})	1
Road: (ie α^{road})	4.5
River: (ie α^{river})	3
Coast: (ie α^{coast})	2.25
Observations	7329
R-squared	0.84

Note: Regressions include salt type x year, and salt type x destination fixed effects, and a salt type x destination trend. OLS standard errors clustered at the destination district level.

Trade Costs: Baseline Results

$$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \delta \ln LCR(\mathbf{R}_t; \boldsymbol{\alpha})_{odt} + \varepsilon^o_{dt}$$

Dependent variable: log destination salt price	OLS (1)	NLS (2)
Log distance to source along lowest-cost route (ie LCR($\mathbf{R_t}$, α))	0.135	0.247
	(0.038)***	(0.063)***
Mode-wise relative marginal costs		
Rail: (ie α^{rail})	1	1
Road: (ie α^{road})	4.5	7.88***
River: (ie α^{river})	3	3.82***
Coast: (ie α^{coast})	2.25	3.94*
Observations	7329	7329
R-squared	0.84	0.97

Note: Regressions include salt type x year, and salt type x destination fixed effects, and a salt type x destination trend. OLS standard errors clustered at the destination district level.

Trade Costs: Extensions

$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \rho RAIL_{odt} + \varepsilon^o_{dt}$

Dependent variable: log destination salt price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Railroad from source to to destination	-0.112 (0.046)***			
Observations	7,329			
R-squared	0.84			

Note: Regressions include salt type x year and salt type x destination fixed effects, and a salt type x destination trend. Column 3 also contains bilateral district pair fixed effects. OLS standard errors clustered at the destination district level.

Trade Costs: Extensions

$$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \rho RAIL_{odt} + \varepsilon^o_{dt}$$

Dependent variable: log destination salt price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Railroad from source to to destination	-0.112	-0.009		
	(0.046)***	(0.041)		
		camels, elephants, carts and inland boats		
Observations	7,329	5,176		
R-squared	0.84	0.73		

Note: Regressions include salt type x year and salt type x destination fixed effects, and a salt type x destination trend. Column 3 also contains bilateral district pair fixed effects. OLS standard errors clustered at the destination district level.

Trade Costs: Extensions

$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \rho RAIL_{odt} + \varepsilon^o_{dt}$

Dependent variable: log destination salt price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Railroad from source to to destination	-0.112	-0.009	-0.046	
	(0.046)***	(0.041)	(0.009)***	
Observations	7,329	5,176	631,451	
R-squared	0.84	0.73	0.76	

If conduct salt regression on ALL bilateral market pair comparisons

Note: Regressions include salt type x year and salt type x destination fixed effects, and a salt type x destination trend. Column 3 also contains bilateral district pair fixed effects. OLS standard errors clustered at the destination district level.

Trade Costs: Extensions

$\ln p^o_{dt} = \beta^o_{ot} + \beta^o_{od} + \phi^o_{od} t + \rho RAIL_{odt} + \varepsilon^o_{dt}$

Dependent variable: log destination salt price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Railroad from source to to destination	-0.112	-0.009	-0.046	-0.024
	(0.046)***	(0.041)	(0.009)***	(0.019)
Observations	7,329	5,176	631,451	9,184,552
R-squared	0.84	0.73	0.76	0.81

If conduct same regression on ALL bilateral market pair comparisons for 17 ag. goods

Note: Regressions include salt type x year and salt type x destination fixed effects, and a salt type x destination trend. Column 3 also contains bilateral district pair fixed effects. OLS standard errors clustered at the destination district level.

Trade Costs: Robustness Checks

- Insignificant changes when allowing for:
  - Divergent technological progress and/or input costs (allow α to change over time)
  - Cost for changing railroad gauge

- 'Out-of-sample' test for free arbitrage violations: How often is $\ln p_{it}^k - \ln p_{jt}^k > \widehat{\delta} \ln LCR(\mathbf{R}_t; \widehat{\alpha})_{ijt}$?
  - 2.8 % of (non-source) pairs for salt
  - 4.8 % of all pairs for 17 agricultural goods

▸ Ad valorem; Congestion ▸ Placebo

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	...reduce price responsiveness?	Yes: to ≈ 0	Model evaluation
4	...raise real income level?	Yes	
5	...reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Railroads and Trade Flows: Summary I

$$\ln \frac{X_{od}^k}{X_d^k} = \ln \lambda^k + \ln A_o^k - \theta_k \ln r_o - \theta_k \ln T_{od}^k + \theta_k \ln p_d^k$$

- Suggests specification (based on earlier proxy for T_{od}^k):

$$\begin{aligned} \ln X_{odt}^k = \beta_{ot}^k &+ \beta_{dt}^k + \beta_{od}^k + \phi_{od}^k t \\ &- \theta_k \widehat{\delta} \ln LCR(\mathbf{R}_t; \widehat{\boldsymbol{\alpha}})_{odt} + \varepsilon_{odt}^k \end{aligned}$$

- Data: 6 million observations on trade flows
  - Geography: 45 Indian 'trade blocks', 23 foreign countries
  - Goods: salt, 17 agricultural
  - Modes: Rail, River, Coast (and some Road)

Railroads and Trade Flows: Summary II

$\ln X_{odt}^k = \beta_{ot}^k + \beta_{dt}^k + \beta_{od}^k + \phi_{od}^k t - \theta_k \widehat{\delta} \ln LCR(\mathbf{R}_t; \widehat{\alpha})_{odt} + \varepsilon_{odt}^k$

- Step 1: Goal is to estimate θ_k
  - Separate regression on each k
  - $\Rightarrow$ average $\widehat{\theta}_k = 3.8$

Railroads and Trade Flows: Summary II

$\ln X^k_{odt} = \beta^k_{ot} + \beta^k_{dt} + \beta^k_{od} + \phi^k_{od} t - \theta_k \widehat{\delta} \ln LCR(\mathbf{R}_t; \widehat{\alpha})_{odt} + \varepsilon^k_{odt}$

- Step 1: Goal is to estimate θ_k
  - Separate regression on each k
  - $\Rightarrow$ average $\widehat{\theta}_k = 3.8$
- Step 2: Goal is to estimate A^k_{ot}
  - Assume: $A^k_{ot} = \gamma_{ot} + \gamma^k_o + \gamma^k_t + \kappa RAIN^k_{ot} + \varepsilon^k_{ot}$
  - $\Rightarrow \widehat{\beta}^k_{ot} + \widehat{\theta}_k \ln r_{ot} = \gamma_{ot} + \gamma^k_o + \gamma^k_t + \kappa RAIN^k_{ot} + \varepsilon^k_{ot}$
  - $RAIN^k_{ot}$: crop k-specific rainfall, from daily rainfall (3614 gauges) and *Crop Calendar* Rain gauges
  - $\Rightarrow \widehat{\kappa} = 0.441$ (0.082)

More Trade Flows
Step 3: Price Responsiveness

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	...reduce price responsiveness?	Yes: to ≈ 0	Model evaluation
4	...raise real income level?	Yes	
5	...reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Railroads and Real Income Levels

- Prediction 4: $\frac{d(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}})}{dT^k_{odt}} < 0$
- Suggests linear approximation:

$$\ln(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}) = \beta_o + \beta_t + \gamma RAIL_{ot} + \varepsilon_{ot}$$

- Data on real agricultural income per acre:
  - $Y_{ot} = \sum_k p^k_{ot} q^k_{ot}$, 17 agricultural crops (ignores: savings, taxes/transfers, intermediate inputs, income from other sectors, income inequality)
  - $\widetilde{P}_{ot}$ = (chain-weighted) Fisher ideal price index, 17 agricultural crops (ignores: other costs of living, gains from new varieties)

Real Income Levels: Reduced-form Results

$\ln(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}) = \beta_o + \beta_t + \gamma RAIL_{ot} + \varepsilon_{ot}$

Dependent variable: log real agricultural income	OLS (1)	OLS (2)
Railroad in district	0.165	
	(0.056)***	
Railroad in neighboring district		
Observations	14,340	
R-squared	0.744	

Note: Regressions include district and year fixed effects. OLS standard errors clustered at the district level.

Alternative RAIL measures Robustness

Real Income Levels: Reduced-form Results

$$\ln\left(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}\right) = \beta_o + \beta_t + \gamma RAIL_{ot} + \phi \frac{1}{N_o} \sum_{d \in N_o} RAIL_{dt} + \varepsilon_{ot}$$

Dependent variable: log real agricultural income	OLS (1)	OLS (2)
Railroad in district	0.165	0.182
	(0.056)***	(0.071)***
Railroad in neighboring district		-0.042
		(0.020)**
Observations	14,340	14,340
R-squared	0.744	0.758

Note: Regressions include district and year fixed effects. OLS standard errors clustered at the district level.

▸ Alternative RAIL measures ▸ Robustness

Robustness Checks

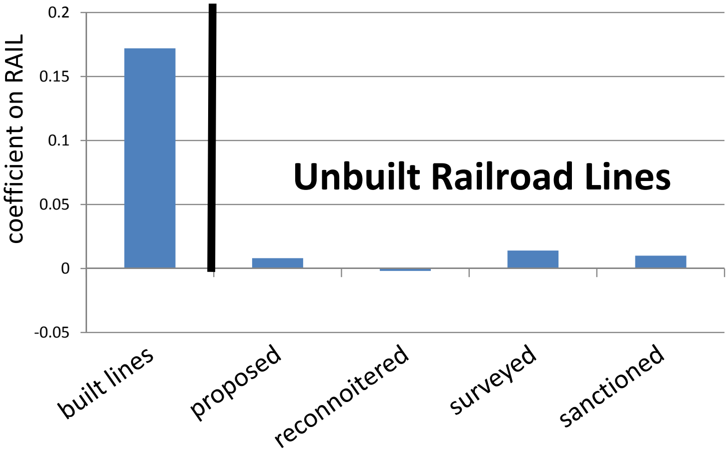
1. 4 Placebo checks [no spurious ‘impacts’]
    - Over 40,000 km of planned lines that were not built for 4 different reasons
2. Instrumental variable [similar to OLS]
    - 1880 Famine Commission: rainfall in 1876-78 predicts railroad construction post-1884
3. Bounds check [tight bounds]
    - Lines explicitly labeled as ‘commercial’, ‘military’ or ‘redistributive’ display similar effects

Robustness Checks

1. 4 Placebo checks [no spurious 'impacts']
   - Over 40,000 km of planned lines that were not built for 4 different reasons
2. Instrumental variable [similar to OLS]
   - 1880 Famine Commission: rainfall in 1876-78 predicts railroad construction post-1884
3. Bounds check [tight bounds]
   - Lines explicitly labeled as 'commercial', 'military' or 'redistributive' display similar effects

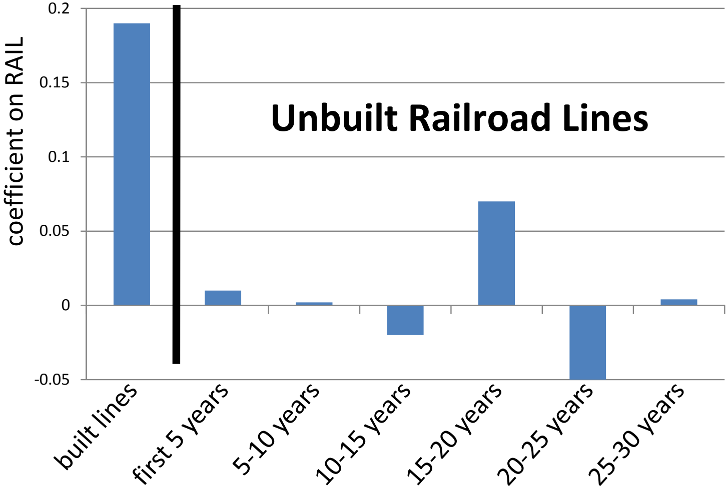
‘Placebo’ I: 4-Stage Planning Hierarchy

14,000 km: Lines reached increasingly costly stages but then abandoned



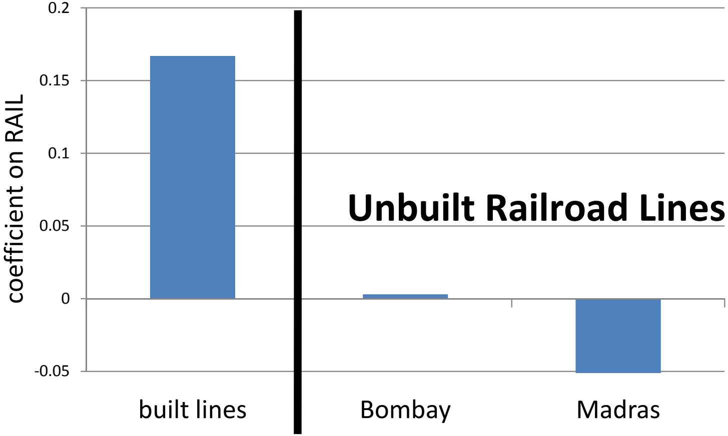
‘Placebo’ II: 1869 Lawrence Plan

12,000 km: Grand 30-year plan scrapped *en masse* by successor



‘Placebo’ III: Chambers of Commerce Plan

7,500 km: Bombay and Madras Chambers submit (commercially attractive) plan



‘Placebo’ IV: Major Kennedy 1853 Plan

9,000 km: Chief Engineer’s cheapest way to connect capitals

Dependent variable: log real agricultural income	OLS (1)	OLS (2)
Railroad in district	0.182	0.188
	(0.071)***	(0.075)**
(Kennedy high-priority line) x trend		0.0005
		(0.038)
(Kennedy low-priority line) x trend		-0.001
		(0.026)
Observations	14,340	14,340
R-squared	0.758	0.770

Note: Regressions control for neighboring district railroad access and include district and year fixed effects. OLS standard errors clustered at the district level.

Robustness Checks

1. 4 Placebo checks [no spurious ‘impacts’]
   - Over 40,000 km of planned lines that were not built for 4 different reasons
2. Instrumental variable [similar to OLS]
   - 1880 Famine Commission: rainfall in 1876-78 predicts railroad construction post-1884
3. Bounds check [tight bounds]
   - Lines explicitly labeled as ‘commercial’, ‘military’ or ‘redistributive’ display similar effects

Instrumental Variable

- 1876-78 famine led to 1880 Famine Commission:
  - 1880 Commission unique in recommending railroads
- Instrumental variable:
  - Rainfall anomalies in 1876-78 agricultural years predict railroad construction post-1884
  - Control for contemporaneous and lagged rain
- Falsification:
  - Does rainfall in other "famine" (Commission) years predict railroads? No.
  - Does rainfall in other "famine" (Commission) years correlate with real income? No.

Instrumental Variable Results

Dependent variable:	Railroad in district OLS (1)	Log real ag income IV (2)
(Rainfall deviation in 1876-78) x (post-1884 indicator)	-0.044 (0.018)***	
Rainfall in district	0.013 (0.089)	1.104 (0.461)**
Rainfall in district (lagged 1 year)	-0.003 (0.048)	0.254 (0.168)
(Rainfall in "famine" year) x (post-"famine" year indicator)	0.006 (0.021)	0.011 (0.031)
Railroad in district		0.197 (0.086)**
Observations	14,340	14,340
R-squared	0.65	0.74

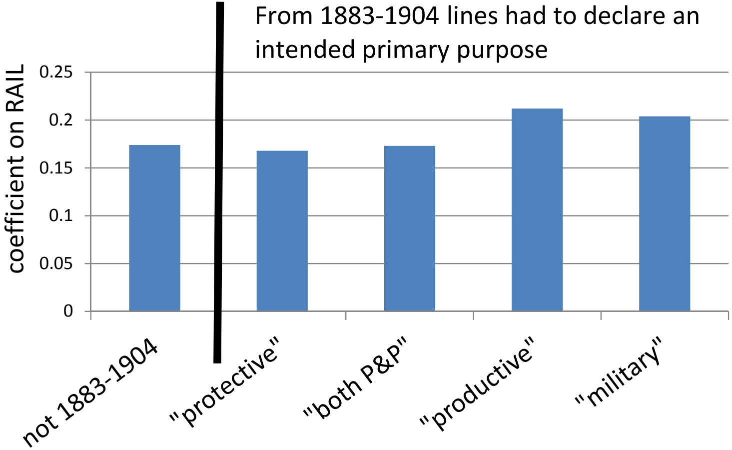
Note: Regressions include district and year fixed effects, and control for rainfall of 2 lagged and 3 lagged years, and neighboring district railroad access. OLS standard errors clustered at the district level.

Robustness Checks

1. 4 Placebo checks [no spurious ‘impacts’]
   - Over 40,000 km of planned lines that were not built for 4 different reasons
2. Instrumental variable [similar to OLS]
   - 1880 Famine Commission: rainfall in 1876-78 predicts railroad construction post-1884
3. Bounds check [tight bounds]
   - Lines explicitly labeled as ‘commercial’, ‘military’ or ‘redistributive’ display similar effects

Bounds Check

$$\ln(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}) = \beta_o + \beta_t + \sum_j \gamma^j PURPOSE^j \times RAIL_{ot} + \phi \frac{1}{N_o} \sum_{d \in N_o} RAIL_{dt} + \varepsilon_{ot}$$



Real Income: Extensions

- Consistent with model's predictions:
  - Bilateral (Krugman) specialization index rises
  - Real income volatility falls ▸ Volatility
- Railroads and demographic change:
  - Mortality rate: 3 % drop
  - Fertility rate: 4 % rise
  - Migration: no change
  - Population: 6 % rise
  - Real agricultural income per capita: 10 % rise
  - Real rural wage: 8 % rise
  - 'Real' urban wage: no change

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	...reduce price responsiveness?	Yes: to ≈ 0	Model evaluation
4	...raise real income level?	Yes	
5	...reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Real Income Gains: Gains from Trade?

- Prediction 6: Autarkiness (π^k_{oot}) is a sufficient statistic for the impact of railroads on real income:

$$\ln\left(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}\right) = \Omega + \sum_k \frac{\mu_k}{\theta_k} \ln A^k_{ot} - \sum_k \frac{\mu_k}{\theta_k} \ln \pi^k_{oot}$$

- Use this to compare reduced-form real income estimates (Step 4) to model predictions:

$$\ln\left(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}\right) = +\rho_1 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k} \widehat{\kappa} RAIN^k_{ot} + \rho_2 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k} \ln \pi(\widehat{\boldsymbol{\Theta}}, \mathbf{Z}_t)^k_{oot}$$
$$+ \alpha_o + \beta_t + \gamma RAIL_{ot} + \phi \frac{1}{N_o} \sum_{d \in N_o} RAIL_{dt} + \varepsilon_{ot}$$

Real Income: Gains from Trade?

$$\ln\left(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}\right) = \gamma RAIL_{ot} + \frac{1}{N_o}\sum_{d\in N_o} RAIL_{dt} + \rho_1 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k}\widehat{\kappa} RAIN_{ot}^k$$

Dep. var: log real agricultural income	OLS	OLS
Railroad in district	0.182	
	(0.071)***	
Railroad in neighboring district	-0.042	
	(0.020)**	
Rainfall in district		
"Autarkiness" measure (computed in model)		
Observations	14,340	
R-squared	0.744	

Note: Regressions include district and year fixed effects. OLS standard errors clustered at the district level.

Real Income: Gains from Trade?

$$\ln\left(\frac{Y_{ot}}{L_{ot}\widetilde{P}_{ot}}\right) = \gamma RAIL_{ot} + \frac{1}{N_o}\sum_{d\in N_o} RAIL_{dt} + \rho_1 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k}\widehat{\kappa} RAIN_{ot}^k + \rho_2 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k} \ln \widehat{\pi}_{oot}^k$$

Dep. var: log real agricultural income	OLS	OLS
Railroad in district	0.182	0.021
	(0.071)***	(0.096)
Railroad in neighboring district	-0.042	0.003
	(0.020)**	(0.041)
Rainfall in district		1.044
		(0.476)**
"Autarkiness" measure (computed in model)		-0.942
		(0.152)***
Observations	14,340	14,340
R-squared	0.744	0.788

Note: Regressions include district and year fixed effects. OLS standard errors clustered at the district level.

Conclusion

1. Railroads improved the trading environment in India
   - Trade costs (and price gaps) fell
   - Trade flows rose
   - Price responsiveness fell
2. Railroads raised real incomes in India
   - Real income volatility fell too
3. Welfare gains from railroads are well accounted for by a Ricardian model of trade
   - Suggests that static gains from trade were important economic mechanism behind the benefits of railroads

Equilibrium Prices

- Consumers in d face many potential suppliers of each variety
  - They consume the cheapest: $p_d^k(j) = \min_o\{p_{od}^k(j)\}$

$$p_d^k(j) \sim G_d^k(p) = 1 - \exp\left[-[\sum_{o=1}^{D} A_o^k \left(r_o T_{od}^k\right)^{-\theta_k}]\, p^{\theta_k}\right]$$

- Average price within good k:

$$E[p_d^k(j)] \doteq p_d^k = \lambda_1^k \left[\sum_{o=1}^{D} A_o^k \left(r_o T_{od}^k\right)^{-\theta_k}\right]^{-1/\theta_k}$$

Return

From Theory to Empirics

- Adding time:
  - Exogenous variables (A^k_{ot}, T^k_{odt}) vary over time
  - Stochastic productivities ($z^k_{ot}(j)$) re-drawn (iid) every period
  - Parameters (θ_k, ε_k) fixed over time

◂ Return

Prediction 3: Price Responsiveness

- Recall: $p_d^k = \lambda_1^k \left[\sum_{o=1}^{D} A_o^k \left(r_o T_{od}^k\right)^{-\theta_k}\right]^{-1/\theta_k}$

- Prediction 3: Price responsiveness $(\frac{dp}{dA})$ and trade costs (T) around symmetric equilibrium:

$$\underbrace{\frac{d}{dT_{do}^k}\left(\frac{dp_d^k}{dA_d^k}\right) < 0}_{\text{less own responsiveness}} \qquad \underbrace{\frac{d}{dT_{do}^k}\left(\frac{dp_d^k}{dA_o^k}\right) > 0}_{\text{more `connected' responsiveness}}$$

Return

Prediction 4: Real Income Levels

- Welfare (of representative agent owning unit of land) is equal to real income:

$$V(\mathbf{p}_o, r_o) = \frac{r_o}{\widetilde{P}_o} = \frac{Y_o}{L_o \widetilde{P}_o}$$

- Prediction 4: Real income ($\frac{Y}{L\widetilde{P}}$) and trade costs (T) around a symmetric equilibrium:

$$\underbrace{\frac{d\left(\frac{r_o}{\widetilde{P}_o}\right)}{dT^k_{od}} < 0}_{\text{own railroads good}} \qquad \underbrace{\frac{d\left(\frac{r_o}{\widetilde{P}_o}\right)}{dT^k_{jd}} > 0 \quad j \neq o}_{\text{others' railroads bad}}$$

Return

Prediction 5: Real Income Volatility

- Prediction 5: Real income responsiveness $\left(\frac{d(\frac{r}{\widetilde{P}})}{dA}\right)$ and trade costs (T) around a symmetric equilibrium:

$$\frac{d}{dT_{od}^{k}}\left(\frac{d(\frac{r_o}{\widetilde{P}_o})}{dA_o^k}\right) > 0$$

- If productivity (A_o^k) is stochastic, then less responsiveness means less volatility

◂ Return

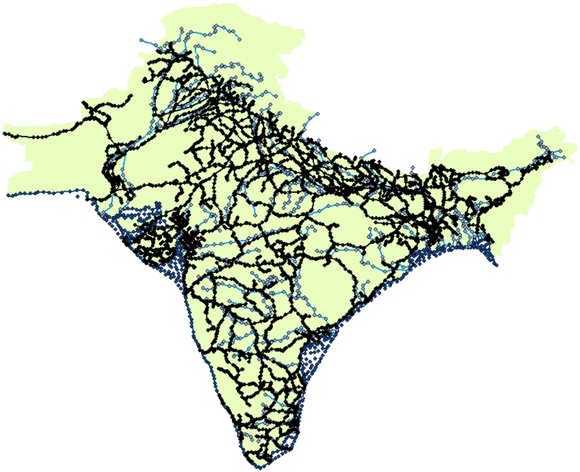
Transport system as a Network

Input: The transportation system (in 1930)

◂ Return

Transport system as a Network

Output: Network representation of transportation system (in 1930)



◂ Return

Trade Costs: Robustness Checks

Ad valorem specification, demand effects, congestion

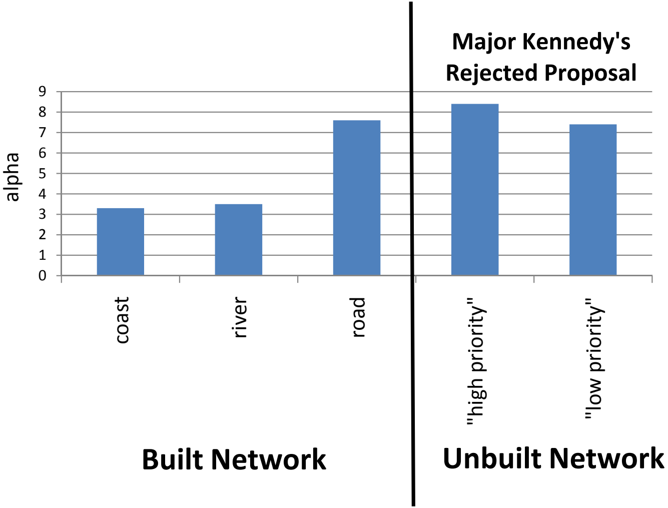
Dependent variable: log destination salt price	NLS (1)	NLS (2)	NLS (3)
Log effective distance to source along LCR	0.247	0.204	0.259
	(0.063)***	(0.076)***	(0.071)***
(Log eff. dist. to source along LCR) x (Excise tax at source)		0.0184	
		(0.040)	
Rainfall at destination			0.013
			(0.042)
Rainfall along source-destination route			-0.003
			(0.081)
Observations	7329	7329	7329
R-squared	0.97	0.98	0.98

Note: Regressions include salt type x year and salt type x destination fixed effects, and a salt type x destination trend. OLS standard errors clustered at the destination district level.

Return

Trade Costs: Major Kennedy's Placebo

Kennedy's 23,000 km prposal. (Recall: $\alpha^{rail} = 1$, for built lines)



Return

Trade Flows: Reduced-form specification

- Prediction 2: $X_{od}^k = \lambda_3^k A_o^k (r_o T_{od}^k)^{-\theta_k} (p_d^k)^{\theta_k} X_d^k$
- Suggests empirical specification:

$$\ln X_{odt}^k = \beta_{ot}^k + \beta_{dt}^k + \beta_{od}^k + \phi_{od}^k t + \rho_1 LCR_{odt} + \rho_2 G^k LCR_{odt} + \varepsilon_{odt}^k$$

- G^k = good-specific characteristics: weight per-unit value (1880), freight class (1880)

◂ Return

Trade Flows: Reduced-form results

$$\ln X_{odt}^k = \beta_{ot}^k + \beta_{dt}^k + \beta_{od}^k + \phi_{od}^k t + \rho_1 LCR_{odt} + \rho_2 G^k LCR_{odt} + \varepsilon_{odt}^k$$

Dependent variable: log value of exports	OLS (1)	OLS (2)	OLS (3)
Fraction of origin-destination districts connected by railroad	1.482 (0.395)***		
Log effective distance to source along lowest-cost route		-1.303 (0.210)***	-1.284 (0.441)***
(Log eff. distance to source along LCR) x (Weight per unit value of good)			-0.054 (0.048)
(Log eff. distance to source along LCR) x (Different freight class from salt)			0.031 (0.056)
Observations	6,581,327	6,581,327	6,581,327
R-squared	0.943	0.963	0.964

Note: Regressions include origin trade block x year x commodity, destination trade block x year x commodity, and origin trade block x destination trade block x commodity fixed effects and an origin trade block x destination trade block x commodity trend. OLS standard errors clustered at the exporting trade block level.

Return

Trade: Estimating parameters—Step 1

- Estimate (once for each good k):

$$\ln X^k_{odt} = \beta^k_{ot} + \beta^k_{dt} + \beta^k_{od} + \phi^k_{od} t - \theta_k \widehat{\delta} \ln LCR(\mathbf{R}_t; \widehat{\boldsymbol{\alpha}})_{odt} + \varepsilon^k_{odt}$$

Sample	Mean ($\widehat{\theta}_k$)	Std. dev. ($\widehat{\theta}_k$)
all 85 goods	5.2	2.1
17 ag. goods	3.8	1.2
Eaton-Kortum OECD manuf.	8.3	$\{3.60, 12.86\}$

Return

Trade: Estimating parameters—Step 2

- Estimate determinants of (agricultural) productivity:
  - Fixed effect $\widehat{\beta}^k_{ot}$ from previous regression interpreted as:

$$\widehat{\beta}^k_{ot} + \widehat{\theta}_k \ln r_{ot} = \ln A^k_{ot}$$

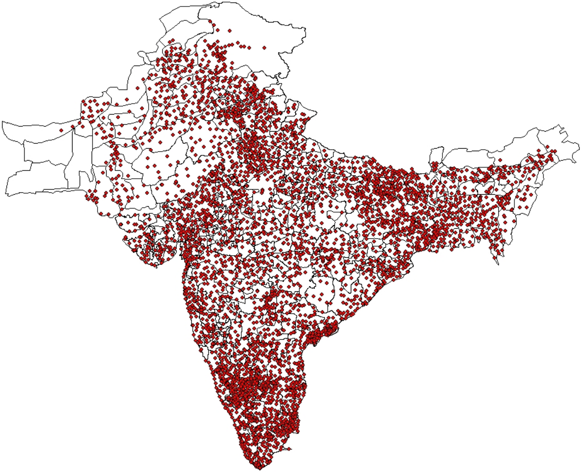
$$\Rightarrow \quad \widehat{\beta}^k_{ot} + \widehat{\theta}_k \ln r_{ot} = \gamma^k_o + \gamma^k_t + \gamma_{ot} + \kappa RAIN^k_{ot} + \varepsilon^k_{ot}$$

- Data:
  - r_{ot} = per acre agricultural output value (17 crops)
  - Crop-specific rainfall from dates in *Crop Calendar*
  - Daily rainfall (3614 rain gauges) Rain gauges
- Result:
  - $\widehat{\kappa} = 0.441 \ (0.082)$

Return

Daily Rainfall Data

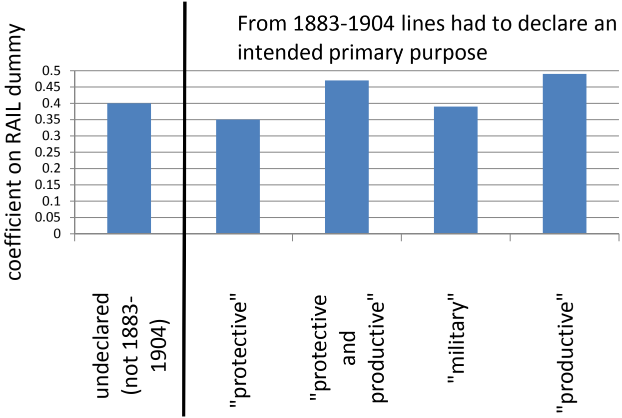
3614 meteorological stations with rain gauges



◂ Return Trade   ◂ Return Prices

Trade Flows: Bounds Check

$$\ln X_{odt}^{k} = \alpha_{ot}^{k} + \beta_{dt}^{k} + \gamma_{od}^{k} + \phi_{od}^{k} t + \sum_j \rho^j TC_{odt} \times PURPOSE^j + \varepsilon_{odt}^{k}$$



Return

Step	Did railroads...	Result	Estimation
1	...reduce trade costs (and price gaps)?	Yes	Trade costs
2	...expand trade?	Yes	Model parameters
3	...reduce price responsiveness?	Yes: to ≈ 0	Model evaluation
4	...raise real income level?	Yes	
5	...reduce real income volatility?	Yes	
6	...promote (static) gains from trade?	Yes: Trade model accounts for 88 % of real income gains	

Prices and Local Rainfall

- Prediction 3: $\frac{d}{dT^k_{dot}} \left| \frac{dp^k_{dt}}{dA^k_{dt}} \right| > 0$
- Suggests linear approximation:

$$\ln p^k_{dt} = \beta^k_d + \beta^k_t + \beta_{dt} + \chi_1 RAIN^k_{dt} + \chi_2 RAIN^k_{dt} \times RAIL_{dt} + \varepsilon^k_{dt}$$

- Data:
  - p^k_{dt} = 239 districts, 17 crops, annually 1861-1930
  - $RAIN^K_{dt}$ = amount of rain over district-crop growing period
  - *Crop Calendar* and daily rain from 3614 gauges

Rain gauges

Return

Price Responsiveness Results

$$\ln p_{dt}^k = \beta_d^k + \beta_t^k + \beta_{dt} + \chi_1 RAIN_{dt}^k + \chi_2 RAIN_{dt}^k \times RAIL_{dt} + \varepsilon_{dt}^k$$

Dependent variable: log price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Local rainfall	-0.256			
	(0.102)**			
(Local rainfall) x (Railroad in district)				
Neighboring district rainfall				
(Neighboring district rainfall) x (Connected to neighbor by rail)				
Observations	73,000			
R-squared	0.89			

Note: Regressions include crop x year, district x year and district x crop fixed effects. OLS standard errors clustered at the district level.

Return Model Evaluation Placebo 1 Placebo 2 Bounds

Price Responsiveness Results

$$\ln p_{dt}^k = \beta_d^k + \beta_t^k + \beta_{dt} + \chi_1 RAIN_{dt}^k + \chi_2 RAIN_{dt}^k \times RAIL_{dt} + \varepsilon_{dt}^k$$

Dependent variable: log price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Local rainfall	-0.256	-0.428		
	(0.102)**	(0.184)***		
(Local rainfall) x (Railroad in district)		0.414		
		(0.195)**		
Neighboring district rainfall				
(Neighboring district rainfall) x (Connected to neighbor by rail)				
Observations	73,000	73,000		
R-squared	0.89	0.89		

Note: Regressions include crop x year, district x year and district x crop fixed effects. OLS standard errors clustered at the district level.

◂ Return ▸ Model Evaluation ▸ Placebo 1 ▸ Placebo 2 ▸ Bounds

Price Responsiveness Results

$$\ln p_{dt}^k = \beta_d^k + \beta_t^k + \beta_{dt} + \chi_1 RAIN_{dt}^k + \chi_2 RAIN_{dt}^k \times RAIL_{dt} + \varepsilon_{dt}^k$$

Dependent variable: log price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Local rainfall	-0.256	-0.428	-0.402	
	(0.102)**	(0.184)***	(0.125)***	
(Local rainfall) x (Railroad in district)		0.414	0.375	
		(0.195)**	(0.184)*	
Neighboring district rainfall			-0.021	
			(0.018)	
(Neighboring district rainfall) x			-0.082	
(Connected to neighbor by rail)			(0.036)**	
Observations	73,000	73,000	73,000	
R-squared	0.89	0.89	0.90	

Note: Regressions include crop x year, district x year and district x crop fixed effects. OLS standard errors clustered at the district level.

◂ Return ▸ Model Evaluation ▸ Placebo 1 ▸ Placebo 2 ▸ Bounds

Price Responsiveness Results

$$\ln p^k_{dt} = \beta^k_d + \beta^k_t + \beta_{dt} + \chi_1 RAIN^k_{dt} + \chi_2 RAIN^k_{dt} \times RAIL_{dt} + \varepsilon^k_{dt}$$

Dependent variable: log price	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Local rainfall	-0.256	-0.428	-0.402	0.004
	(0.102)**	(0.184)***	(0.125)***	(0.035)
(Local rainfall) x (Railroad in district)		0.414	0.375	0.024
		(0.195)**	(0.184)*	(0.120)
Neighboring district rainfall			-0.021	
			(0.018)	
(Neighboring district rainfall) x (Connected to neighbor by rail)			-0.082	
			(0.036)**	
Observations	73,000	73,000	73,000	8,489
R-squared	0.89	0.89	0.90	0.53

Salt

Note: Regressions include crop x year, district x year and district x crop fixed effects. OLS standard errors clustered at the district level.

Return Model Evaluation Placebo 1 Placebo 2 Bounds

Model Validation Using Price Data I

- Recall, prices: $p_d^k = \lambda_1^k \left[\sum_{o=1}^{D} A_o^k (r_o T_{od}^k)^{-\theta_k}\right]^{\frac{-1}{\theta_k}}$
- Have estimates of RHS:
  - $A_{ot}^k = \widehat{\kappa} RAIN_{ot}^k$ and $\widehat{\theta}_k$ from trade flows
  - $\ln T_{odt}^k = \widehat{\delta} \ln LCR(\mathbf{N}_t; \widehat{\alpha})_{odt}$ from salt prices
  - r_{ot}: could use data on this, but compute model prediction instead $\Rightarrow \widehat{r}_{ot}$
  - λ_1^k Contains σ_k, but don't need it
- Include predicted prices in regression to evaluate out-of-equation performance

$$\widehat{p}_{dt}^k = \lambda_1^k \left[\sum_{o=1}^{D} \widehat{A}_{ot}^k (\widehat{r}_{ot} \widehat{T}_{odt}^k)^{-\widehat{\theta}_k}\right]^{\frac{-1}{\widehat{\theta}_k}}$$

Return

Model Evaluation using Price Data II

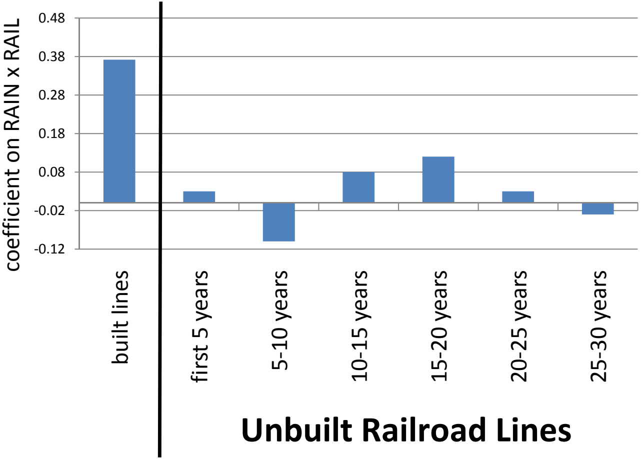
Dependent variable: log price	OLS (1)
Predicted prices	0.913
	(0.189)***
Observations	73,000
R-squared	0.93

Note: Regressions include crop x year, district x year and district x crop fixed effects. OLS standard errors clustered at the district level.

Return

Price Responsiveness: Placebo Checks I

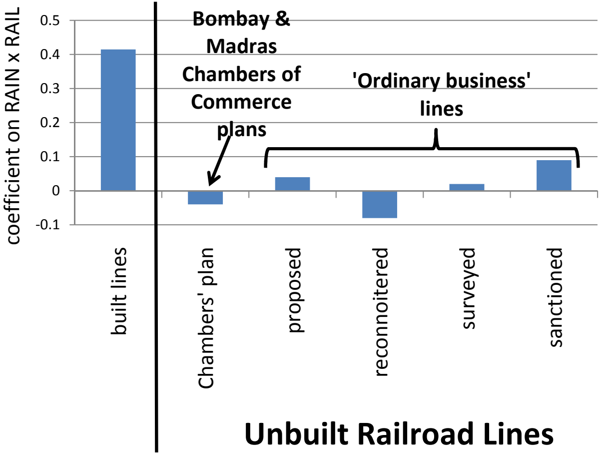
12,000 km Lawrence Plan scrapped *en masse* by successor



Return

Price Responsiveness: Placebo Checks II

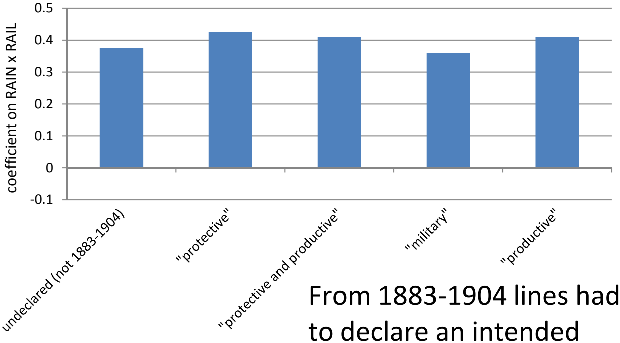
Chambers of Commerce Plans; 4-stage hierarchy



Return

Price Responsiveness: Bounds Check

$$\ln p_{dt}^k = \alpha_d^k + \beta_t^k + \gamma_{dt} + \delta_1 RAIN_{dt}^k + \sum_j PURPOSE^j \gamma^j RAIN_{dt}^k \times RAIL_{dt} + \varepsilon_{dt}^k$$



From 1883-1904 lines had to declare an intended primary purpose

Return

Real Income Levels: Robustness

Dependent variable: log real agricultural income	OLS (1)	OLS (2)	OLS (3)
Railroad in district	0.182	0.197	0.182
	(0.071)***	(0.102)*	(0.095)*
Railroad in neighboring district	-0.042	-0.055	-0.042
	(0.020)**	(0.039)	(0.025)*
District-specific tends	No	Yes	No
Standard errors	Clustered	Clustered	Conley
Observations	14,340	14,340	14,340
R-squared	0.758	0.813	0.758

Note: Regressions include district and year fixed effects. Standard errors clustered at the district level. Conley standard errors calculated using 250 km cut-off.

Return

Alternative Measures of Rail Access

"Average log LCR" $= \frac{1}{N_d} \sum_{d \in N_d} \ln LCR(\mathbf{R}_t; \widehat{\alpha})_{odt}$

Dependent variable: log real agricultural income	OLS (1)	OLS (2)
Railroad in district	0.223	
	(0.091)***	
(Railroad in district) x	-0.064	
(Coastal or riverine district)	(0.036)*	
Average log LCR of district		-0.350
		(0.081)***
Neighbors' average log LCR		0.061
		(0.022)***
Observations	14,340	14,340
R-squared	0.749	0.815

Note: Regressions include district and year fixed effects. Column (1) also controls for neighboring district rail access. OLS standard errors clustered at the district level.

Return

Real Income Volatility

$$\ln\left(\frac{r_{ot}}{\widetilde{P}_{ot}}\right) = \gamma_1 RAIL_{ot} + \rho_1 \sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k} \widehat{\kappa} RAIN_{ot}^k + \gamma_2 RAIL_{ot} \times \left(\sum_k \frac{\widehat{\mu}_k}{\widehat{\theta}_k} \widehat{\kappa} RAIN_{ot}^k\right) + \varepsilon_{ot}$$

Dependent variable: log real agricultural income	OLS (1)	OLS (2)
Railroad in district	0.186	0.252
	(0.085)*	(0.132)*
Rainfall in district	1.248	2.434
	(0.430)***	(0.741)***
(Railroad in district)*(Rainfall in district)		-1.184
		(0.482)***
Observations	14,340	14,340
R-squared	0.767	0.770

Note: Regressions include district, year and province x year fixed effects, and control for neighboring region railroad effects. OLS standard errors clustered at the district level.

Return