Railroads and the Rise of the Factory: Preliminary Evidence for 1850

Jeremy Atack
Michael Haines
Robert A. Margo

1

The Transportation Revolution

- During the 19th Century the US experienced a transportation revolution (Taylor 1951) AND an industrial revolution
- Transportation Revolution = canals, inland waterways, railroads, surface roads
- Industrial Revolution = shift of labor into manufacturing AND rise of factory system
- Factory displaces artisan shop (Sokoloff 1984)

2

Did the TR "cause" the rise of factories?

- Here is one story. Pre-TR: typical mfg firm is very small and may have had significant (local) market power
- Pre-TR: mfg firm faces downward sloping demand, MR = MC to the left of min of U-shaped AC
- Post-TR: mfg firm faces much flatter demand curve and q moves closer to min AC
- Moving down AC towards minimum entails more division of labor (Adam Smith)
- Other implications: % skilled ↓ and TFP ↑
- There are, of course, alternative hypotheses…

3

How to Test Transportation Hypothesis?

- We need sample of mfg establishments without and w/o transportation access and identification strategy.
- Mfg sample is easy part. Atack-Bateman sample of mfg firms.
- Transportation access. New county level database (next slide).
- Identification: difference-in-difference and IV.
- This paper focuses on railroad access. Later versions will extend to canals.

4

Atack, Haines & Margo, "RR and the Rise of the Factory," NBER SI, 2008

New Database: The TR

- Digitized 19th century maps onto county boundary files to generate (thus far) a panel of US counties, 1850-1880 with access to rail.
- Rail access = 1 if railroad passes through county boundary. Previously used (Craig, Palmquist, and Weiss; Beeson, De Jong, Troesken)
- Also presence of a canal, navigable waterway (river) in county, or county abuts Great Lakes or ocean (at the moment, these are fixed as of 1850 county boundaries).
- TR database has FIPS codes so can be linked to other datasets with county level identifiers.

Linked MFG-TR File

- Linked the TR data base to the mfg samples.
- Atack-Bateman: random samples of manufacturing establishments from 1850-80 manuscript censuses of manufacturing (*Historical Methods*, 1999).
- We use national samples for 1850-70. 1880 linked sample needs further work (re-weight for special agent industries).
- Primary outcome for today: Factory = 1 (≥ 16 workers)
- Some additional IV results: % female (↑ more division of labor and lower skill on average), TFP (Value added per unit of labor ↑ holding capital per unit of labor constant

Table 1: Sample Characteristics

Year	1850	1860	1870
% of establishments in counties with rail` access	67.5% {89.0%}	85.3% {93.9%}	95.3% {98.0%}
% factory, establishments in counties with no rail access	3.9%	4.7%	1.3%
% factory, establishments in counties with rail access	10.6%	10.6%	10.7%
Difference, Row 3-Row 2, percentage points	6.7	5.9	9.4
% factory (≥ 16 workers)	8.4% [60.3%]	9.7% [67.1%]	10.3% [71.6%]
N(establishments)	5,492	5,210	4,746

Unit of observation is the manufacturing establishment. County has rail access =1 if railroad passes through county boundary (1850 boundaries). {}: county has rail or water access (canal, river, ocean or Great Lakes border). []: percent of workers employed in factories.

Treatment Effect of Rail Access: I

- Two approaches: (1) difference-in-difference (2) instrumental variables
- DD: Compare outcomes in counties that gain access (in 1850s, 1860s) vs. control sample
- DD analysis uses 1850-1870 mfg samples, with 1850 county boundaries imposed

Atack, Haines & Margo, "RR and the Rise of the Factory," NBER SI, 2008

Table 2: DD Approach

County gains rail access in:	No Controls	No Controls	No Controls, Sample restricted to urban establishments	Urban + 2 digit SIC controls	Urban + 2 digit SIC controls	2-digit SIC controls, Sample restricted to urban establishments	2-digit SIC controls , Sample restricted to Urban establishments with water access
1850s	0.018 (0.017)			0.006 (0.017)			
1860s	0.069* (0.022)			0.056* (0.021)			
After 1850		0.027 (0.016)	0.080 (0.065)		0.015 (0.016)	0.098 (0.059)	0.124* (0.063)
N	15,488	15,488	4,062	15,488	15,488	4,062	3,762

Observations pooled from 1850-70 samples. All regressions include year and county fixed effects. Urban = 1 if establishment is located in town/village/city of population 2,500 or more. Standard errors (in parentheses) are clustered at county level. *Significant at 5 percent level or better. Column 8: sample restricted to urban observations in counties with water access (canal or river passes through county, or ocean or Great Lakes frontage). In this sample, % of establishments in counties with rail access increases from 87.6 percent in 1850 to 98.3 percent in 1870.

Treatment Effect of Rail Access: II

- Instrumental variable: "straight line between two points" (Banerjee, Duflo, and Qian use this IV to estimate effects of rail access in late 20th century China).
- Underlying behavioral model for IV: when connecting two locations by rail, choice favors designs that on average are lower cost to build *ex ante* lower. Straight line is shortest distance between two points.
- First Stage: IV = 1 if county lies on a straight line between two points to be connected
- IV: uses 1850 cross section (so far)

IV estimation: Which Points to Connect?

- Identify all cities/towns with pop > 2,500 in 1820
- Port IV. We take major ports in 1820: Baltimore, Boston, Charleston, New Orleans, New York, Norfolk, Philadelphia, Portland (ME), Savannah
- Draw straight line from each city/town to NEAREST port. If straight line passes through a county, PORT = 1.
- First stage (=1 if county has a RR in 1850) is not bad: PORT coefficient is positive and always significant at 5 percent level (even with state dummy controls although base specification uses census region dummies).
- We tried other straight line instruments but these don't work as well. Example: straight lines between all cities/towns with 5K or 8K pop in 1820 (too saturated. First stage vanishes when regional dummies added).

Table 3: IV for Factory Status, 1850

Dependent Variable	Full Sample	Port Cities Excluded	Port Cities and States with IV = 0 Excluded
Factory = 1	0.189* (0.062)	0.165* (0.058)	0.159* (0.061)
First Stage	0.241* (0.042)	0.247* (0.043)	0.224* (0.043)
Sample mean, factory = 1	0.084	0.076	0.087
N	5,492	5,238	3,986

Dependent variable =1 if number of workers (men + women) >=16. Coefficient estimates are virtually unchanged if one (Sokoloff 1984) is added to the count of workers. Independent variables: urban status (pop>2,500), presence of natural waterway in county (river, ocean access, Great Lakes), presence of canal, 2-digit industry, census region. IV = 1 if county lies on straight line between town/city with 2,500 or more population in 1820 and nearest major port (see previous slide). *significant at 5 percent level or better. Standard errors corrected for clustering at county level.

Atack, Haines & Margo, "RR and the Rise of the Factory," NBER SI, 2008

Table 4: Additional IV Results: I

Dependent Variable	Full Sample	Port Cities Excluded	Port Cities and States with IV = 0 excluded
Percent Female	0.070* (0.028)	0.066* (0.027)	0.054** (0.029)
First Stage	0.241* (0.042)	0.247* (0.043)	0.224* (0.043)
Sample mean, Percent Female	0.054	0.048	0.056
N	5,492	5,238	3,986

Dependent variable = women/(men+ women). See notes to previous table for independent variables.
**significant at 10 percent level.

Table 5: Additional IV Results: II

Dependent Variable	Full Sample	Port Cities Excluded	Port Cities and States with IV=0 excluded
Log (adjusted labor productivity)	0.119 [0.231***] (0.138)	0.061 [0.159] (0.135)	0.033 [0.151] (0.163)
First Stage	0.239* (0.043)	0.246* (0.044)	0.221* (0.044)
Sample mean, log (adjusted labor productivity)	6.111	6.096	5.72
N	5,128	4,881	3,690

Dependent variable = log (value added/adjusted labor input). Value added = value of outputs – value of raw materials. Adjusted labor input = men+0.6*women. Log (capital/adjusted labor input) is included as an independent variable. []: one is added to adjusted labor input (Sokoloff 1984), note that this also changes capital per adjusted worker. ***significant at 15 percent level

Conclusions

- Some (fairly shaky) preliminary evidence that diffusion of railroad was a "cause" of the rise of the factory system.
- Effect of TR on diffusion of the factory NOT taken into account in measuring aggregate impact of TR.
- Exact mechanism not revealed by our analysis. COULD be extent of the market BUT might be something else that the railroad does.
- Example: railroad "smoothes" supply/demand disruptions enabling the firm to operate full year.
- Full year operation → greater likelihood of factory

Conclusions

- More work on why DD does poorly for 1850s.
- Extend analysis to 1820-50 using 1820 Sokoloff sample (also maybe 1832).
- Additional IV based on 1824-1838 transport route surveys by federal civil engineers. Those reported in *American State Papers* give start and end locations, so straight line method can be applied. Surveys made for railroads and canals.
- Current analysis treats canals as pre-determined but canal location was endogenous, too. Transport route surveys may provide an IV for canals.

Atack, Haines & Margo, "RR and the Rise of the Factory," NBER SI, 2008