A Comprehensive GIS Dataset for China's Surface Transport Network

Steven J. Davis, Meijun Qian & Wen Zeng

Transport Infrastructure and Productivity: The China Experience

Steven J. Davis & Meijun Qian

BFI-China Research Conference
University of Chicago
1-2 April 2022

China's (Surface) Transport Network

- **Modern motorways:** At least 2 lanes each way, typical speed limits of 80-120 km/hour. Tolls.
- **National highways**: 2-3 lanes each way in east & coastal areas, 60-80km/hour; mostly 1-2 lanes in middle and west China, speed limits as low as 20km/hour. No tolls.
- **Waterways:** Inland and coastal lines navigable by ships.
- **Railways:** standard & fast for passengers and freight.
- **High-speed railways:** Passenger only, max speeds of 250 km/hour or more.

Our concept of the network includes local access points (e.g., railway stations and on & off ramps for motorways & highways).

What We Do

1. Collect and geocode highly granular annual data on China's surface transport system from 1993 to 2020.

2. Quantify and visualize several aspects of China's transport network and its development over time.

3. Integrate the transport system data with longitudinal data on the location, inputs and outputs of 450,000 manufacturing plants.

4. Describe the empirical relationship of plant-level TFP and its local dispersion to transport network access in the cross section and over time.

What We Do

5. Use quasi-experimental variation to estimate effects of transport network access on
   a) Plant-level TFP, entry and exit
   b) Level and dispersion of TFP within locations (zip codes)

6. Aspirational: Draw on estimated effects to quantify the impact of transport network improvements on local and aggregate TFP, working through:
   a) Entry and exit
   b) Reallocation among continuers
   c) TFP changes at continuers

Our Measurement Contribution

- Partial snapshots of China's transport network in previous studies clearly show its growing scale and complexity over the past century. Examples include Wang et al. (2009), Jiao et al. (2014), Hu et al. (2015) and Jin et al. (2019).
- The growth of high-speed railways and air travel in recent decades have also greatly improved connectivity between Chinese cities: Examples include Lao et al. (2016) and Jiao et al. (2017).
- Relative to earlier work, our GIS dataset is **more granular**, **more comprehensive** in its coverage of surface transport modes, **attentive to route access locations**, and **better suited for tracking year-to-year changes i**n highways, motorways, and railways. We also incorporate **travel speed estimates by modes and routes**.

Basic (But Extraordinary) Facts

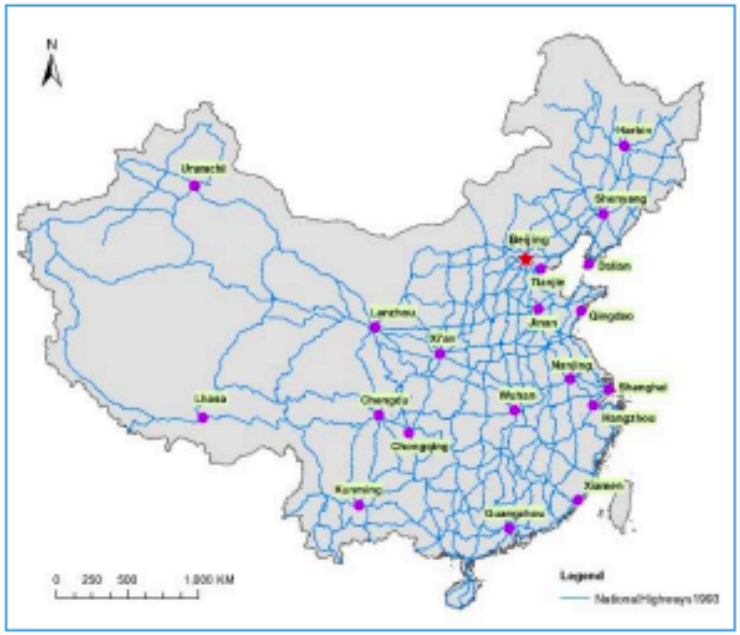
1. The overall length of China's surface transport network more than tripled from 1993 to 2020, reaching 563,000 km.
2. Half the increase took the form of modern motorways and high-speed railways, which were nearly non-existent in the early 1990s.
3. Number of access points to motorways and railways rose ten-fold from 1995 to 2020.
4. Average distance to nearest network access point fell from 13km to 5km.

Basic (But Extraordinary) Facts

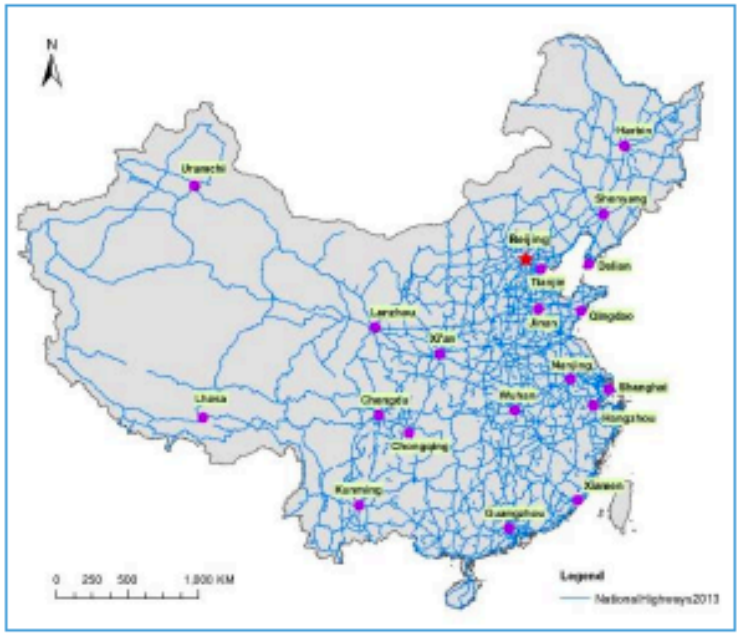
5. Route density nearly doubled, as did beta-index connectivity (i.e., route links per network node, treating zip codes as nodes).
6. Average county-level betweenness centrality nearly tripled.

Note: Straight-line distances to transport routes (widely used in previous research) often deviate greatly from actual distances to route access points. Moreover, we suspect that the magnitude of the deviation varies over time and with population density, output per person, terrain, etc.

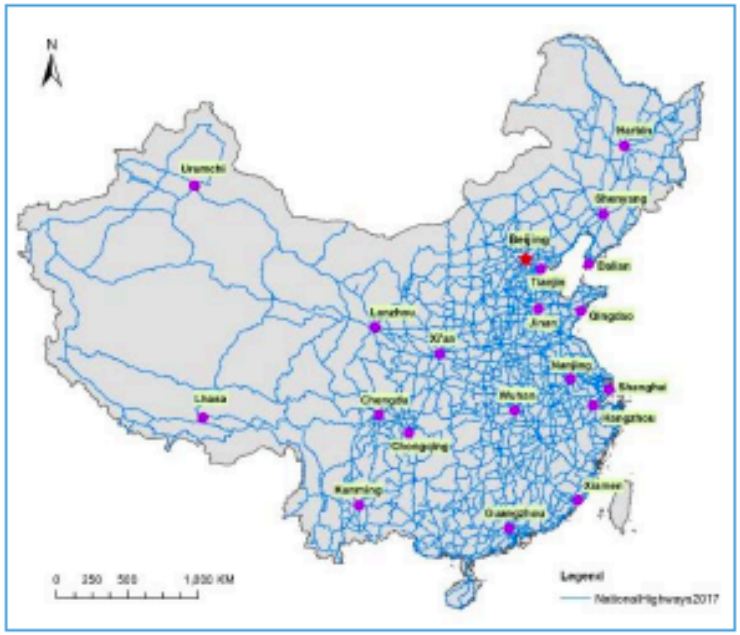
A. Highways, 1993



B. Highways, 2013



C. Highways, 2017

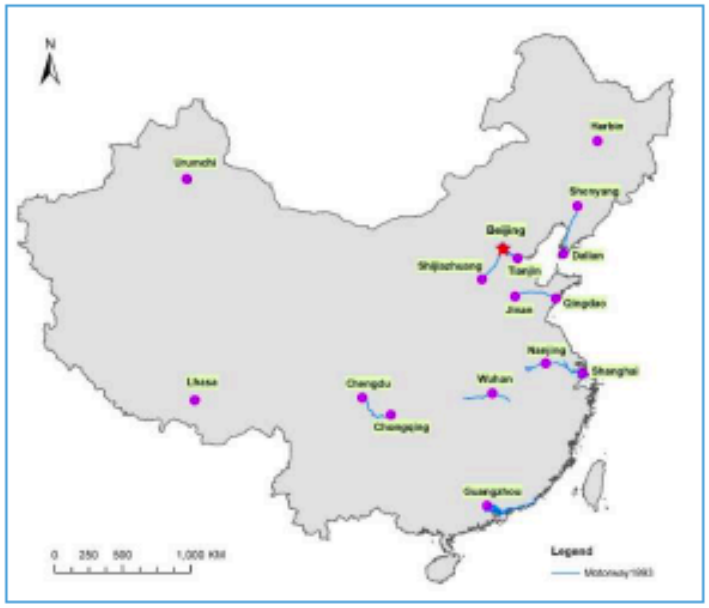


D. Highways, 2020

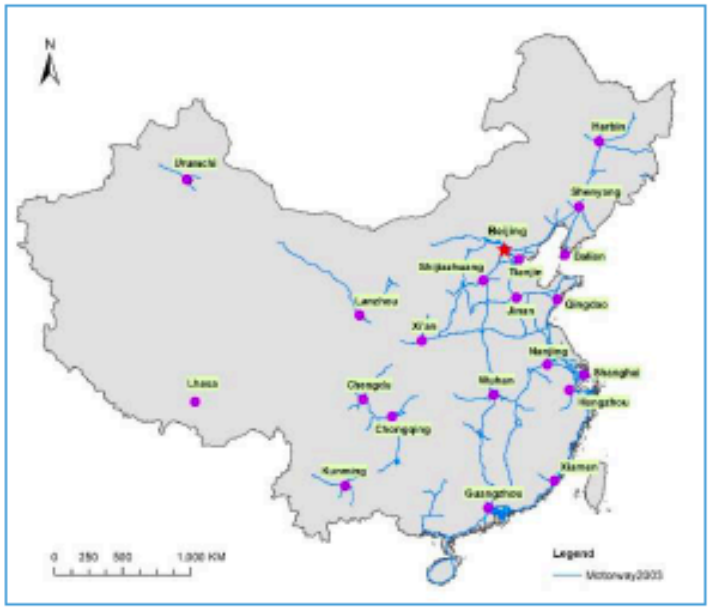
Highways

Extensive system by 1993, but much expansion after 1993.

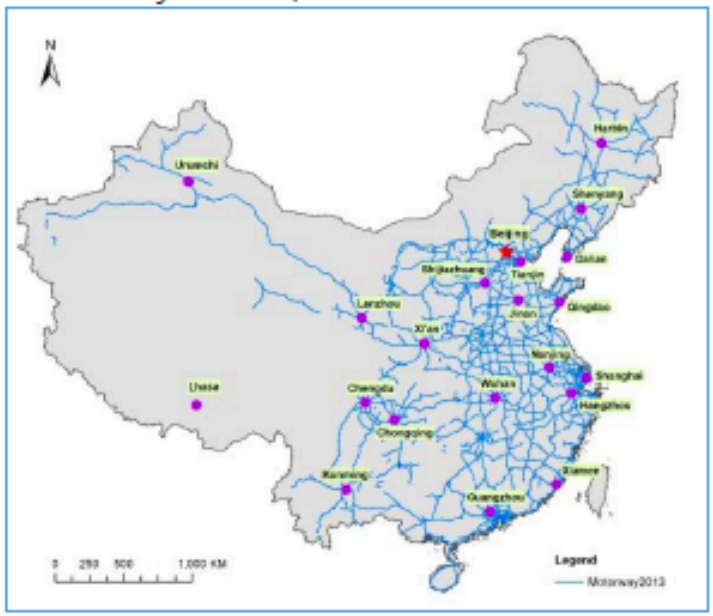
A. Motorway Network, 1993



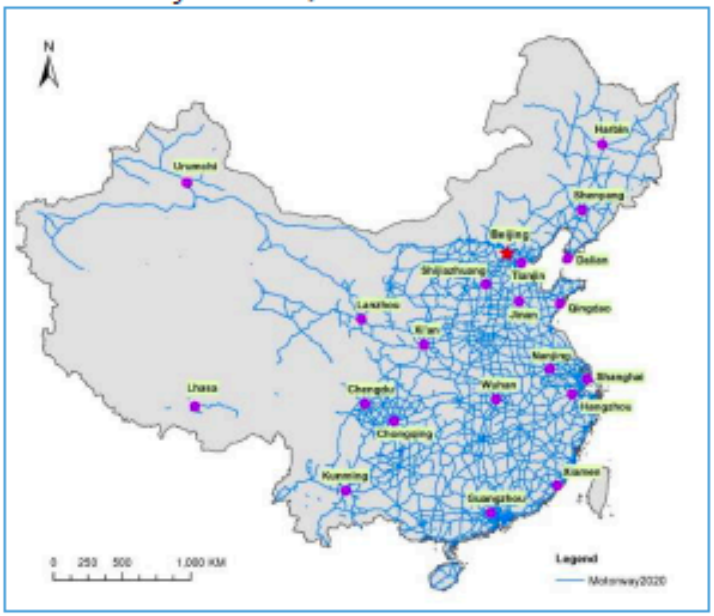
B. Motorway Network, 2003



C. Motorway Network, 2013



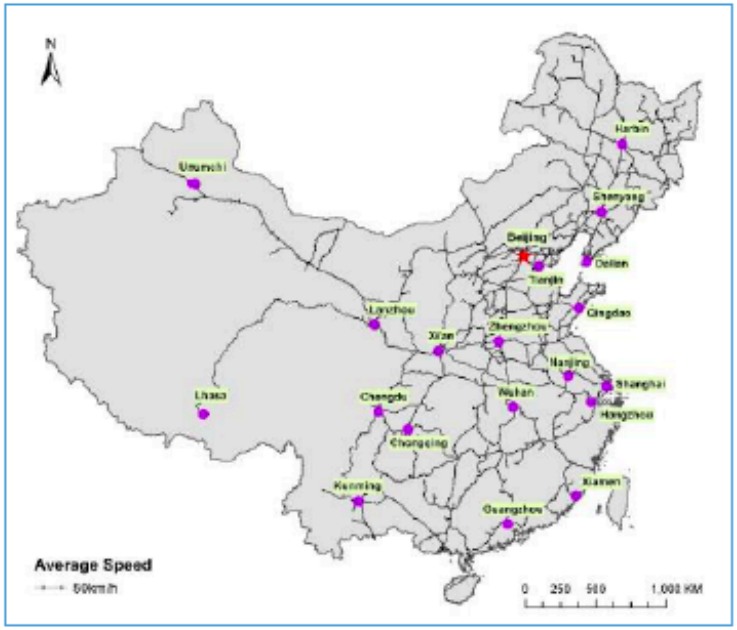
D. Motorway Network, 2020



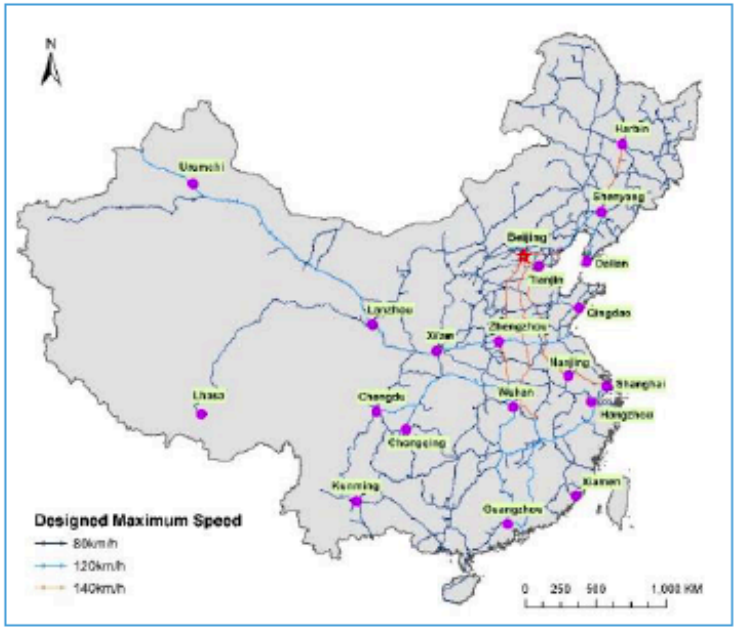
Motorways

Nearly non-existent in 1993, but rapid build out in last 20 years in Eastern half of China.

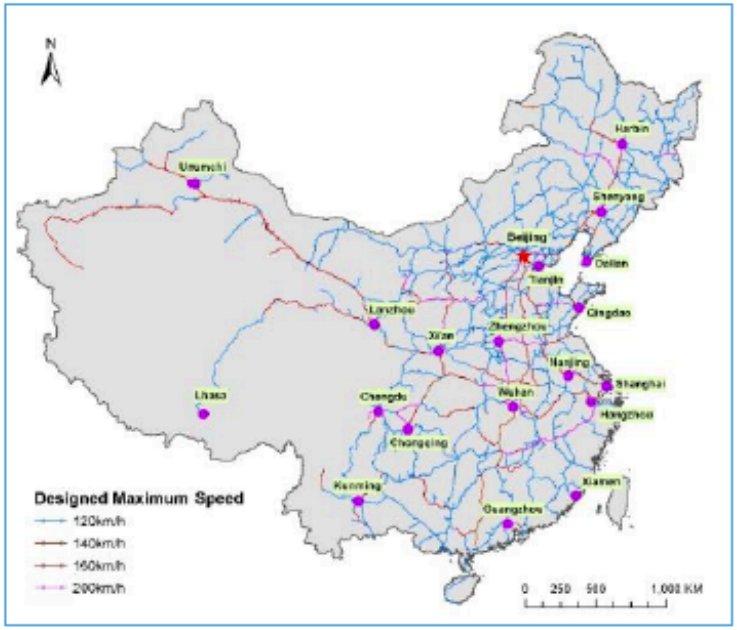
A. Railway Network, 1995



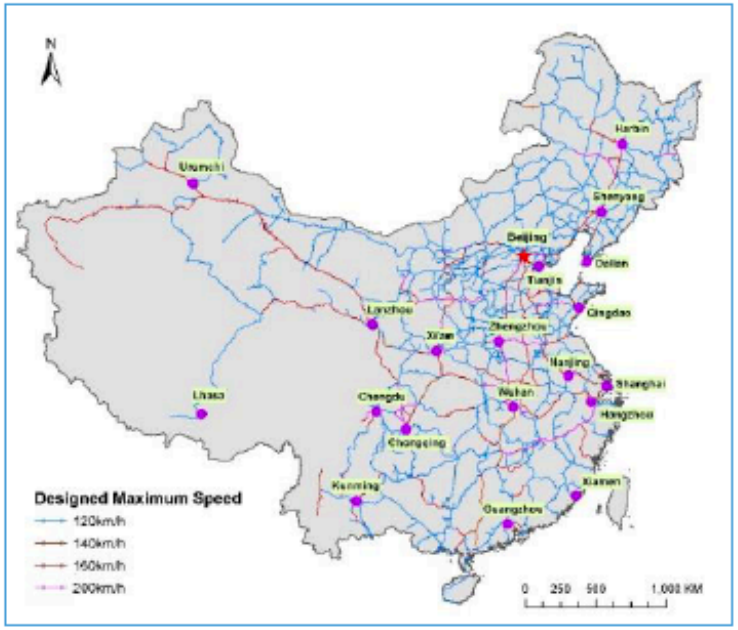
B. Railway Network, 2003



C. Railway Network, 2013



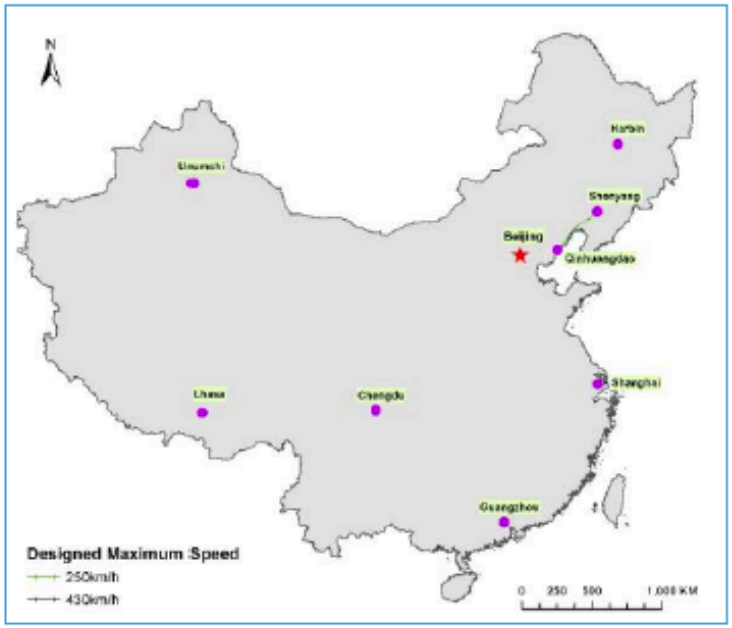
D. Railway Network, 2020



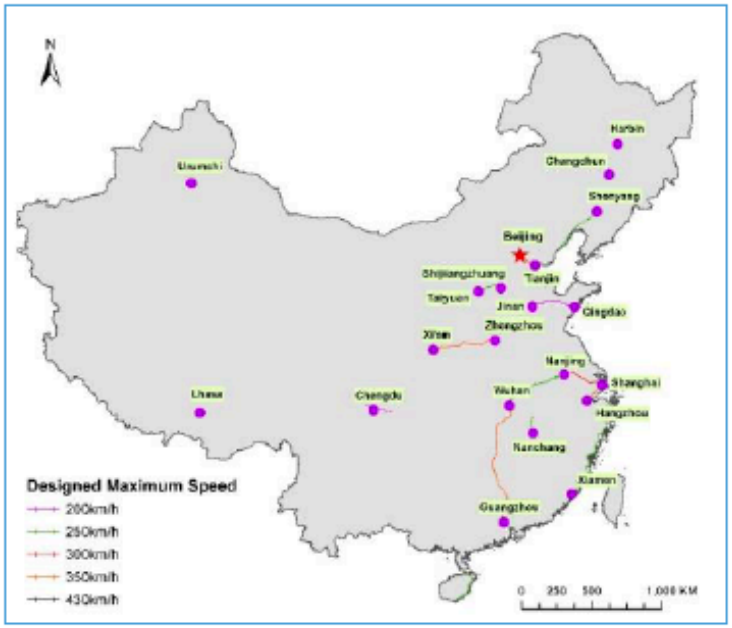
Railways

Large system by 1993, but much expansion and speed gains since 1993.

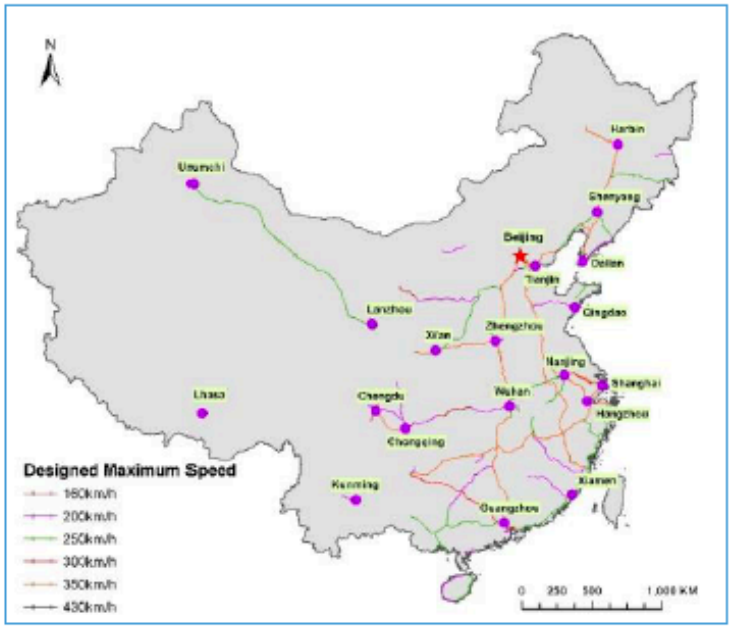
A. High-speed Railway Network, 2003



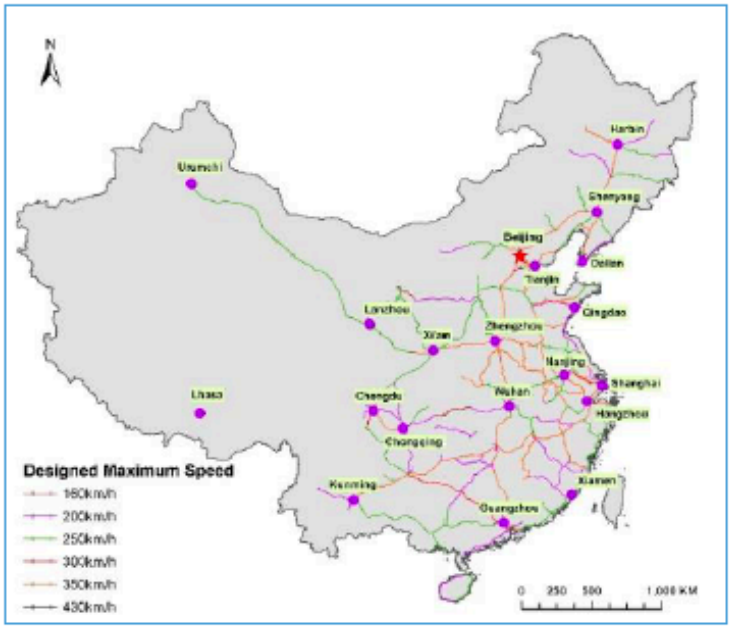
B. High-speed Railway Network, 2010



C. High-speed Railway Network, 2015



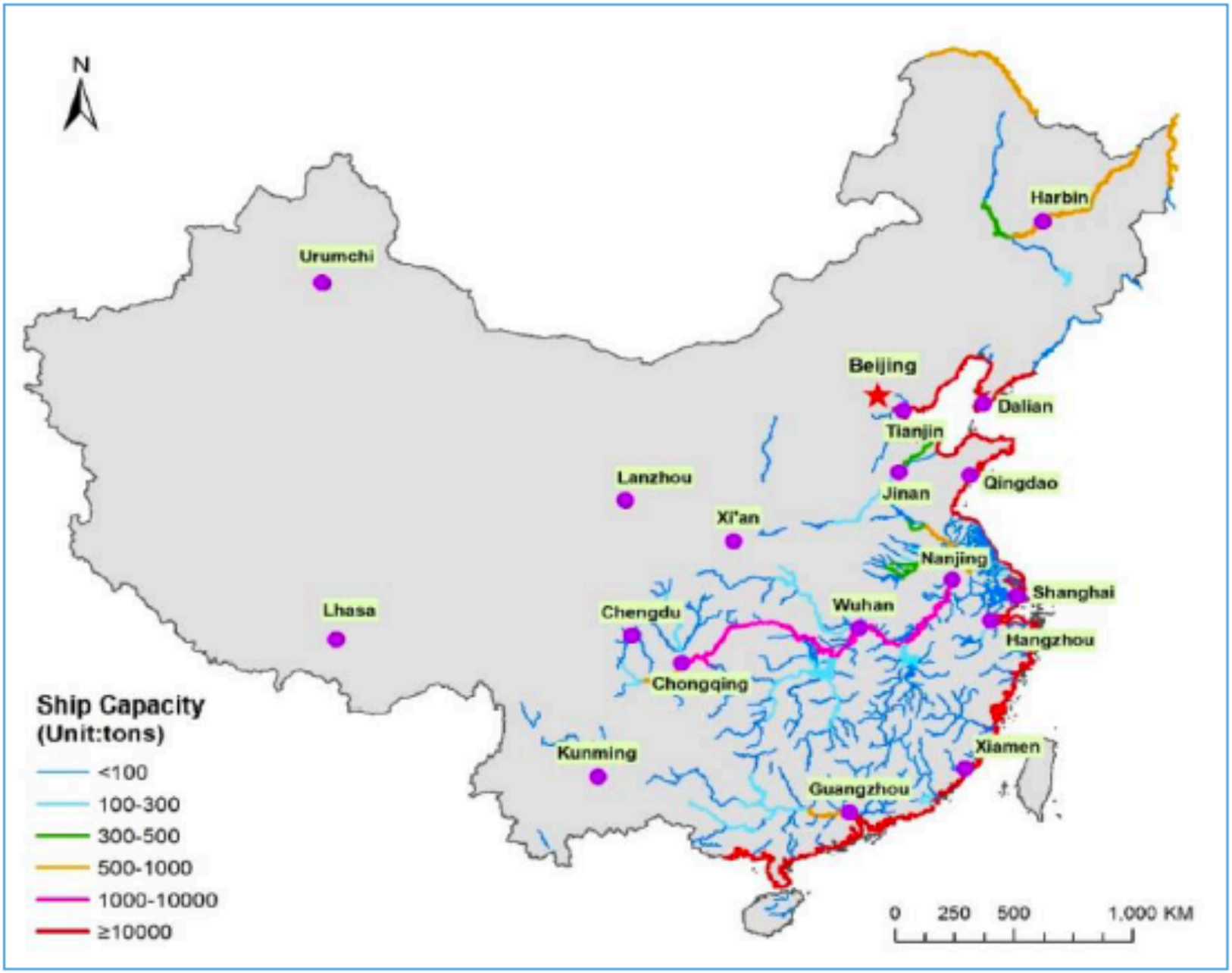
D. High-speed Railway Network, 2020

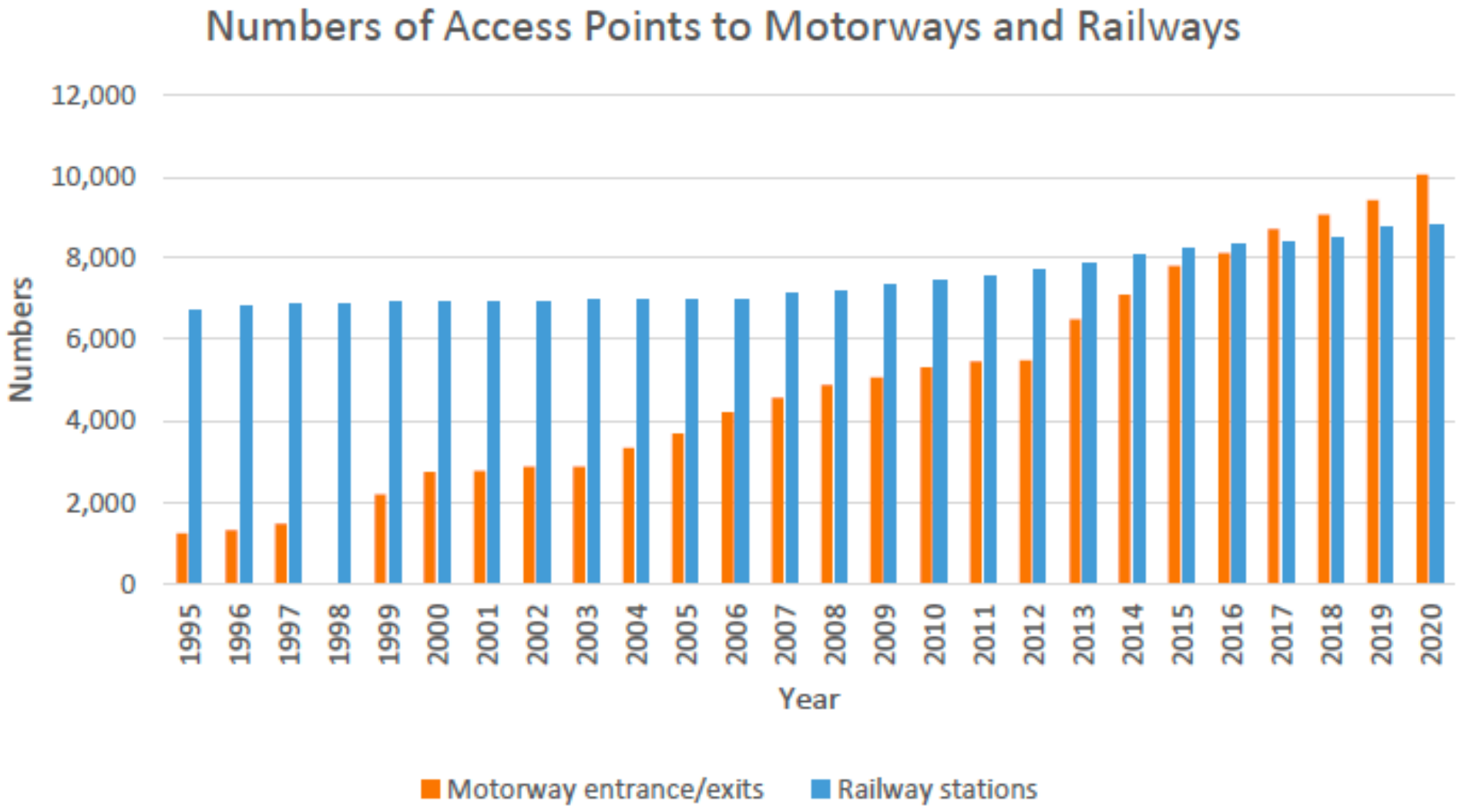


Hi-Speed Railways

Huge growth since 2010, very little before then.

Major Navigable Waterways: 1994




Numbers of Access Points to Motorways and Railways
12,000
10,000
8,000
6,000
4,000
2,000
0
Numbers
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year
Motorway entrance/exits
Railway stations

Effects of China's Transport Network: Previous Works

1. Zheng & Khan (2013), Qin (2017) and Lin (2017) find that high-speed railways facilitate market integration and urban employment, raise nearby house prices, and reduce output along upgraded railway routes.
2. Faber (2014) finds that China's national highway system affects concentration of economic activity and lowers growth in non-targeted peripheral counties.
3. Zhang and Ji (2019) find that railway and road development can raise or lower local output, depending on how it alters competition with other regions.
4. Banerjee et al. (2020) find proximity to railways and waterways has modest positive effects on per capita GDP but not on growth.
5. Baum-Snow et al. (2020) find that better access to national highways raises output and population in regional centers at the expense of hinterland prefectures.

Our Manufacturing Data

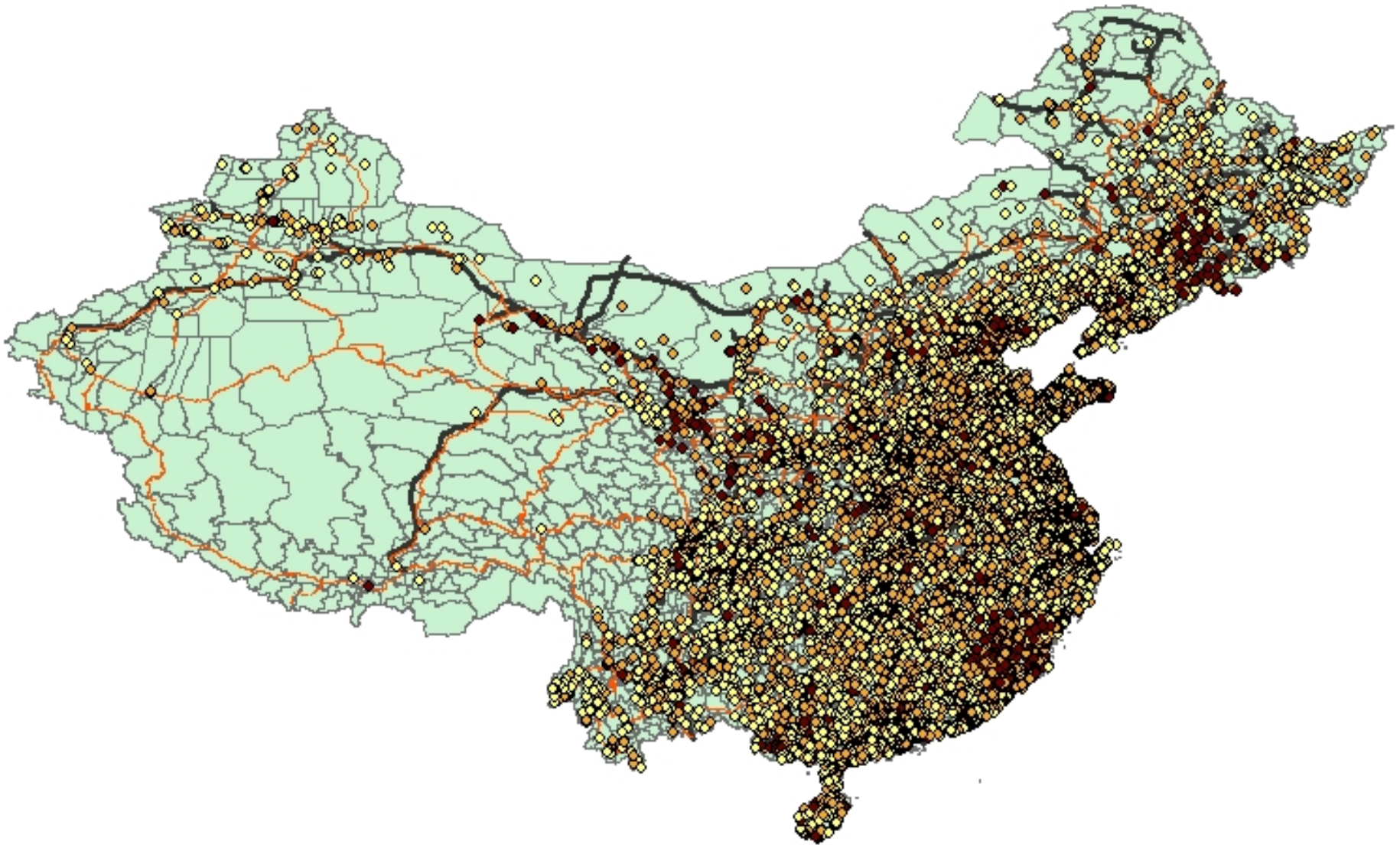
- Plant-level database covers all SOEs and most non-SOE production
  - All non-SOE plants with annual sales > 5 million RMB (about $725K)
  - Age measure lets us distinguish true entry from crossing size threshold
  - 89% of plants report as single-location firms in 1998, 97% in 2007
- Annual data from 1998-2013, but we stop in 2007 due to data quality problems in later years (Gao et al., 2019)
  - We drop plants that appear in only one year
  - We drop plants in 288 counties that each host fewer than 13 plants
  - **After exclusions, our sample covers 450 thousand plants and 2+ million plant-year observations in nearly 2,900 counties**
- Plant location = longitude/latitude of zip code centroid.
- We compute distance of each plant to nearby access points for each transportation mode

Potential Effects on Productivity

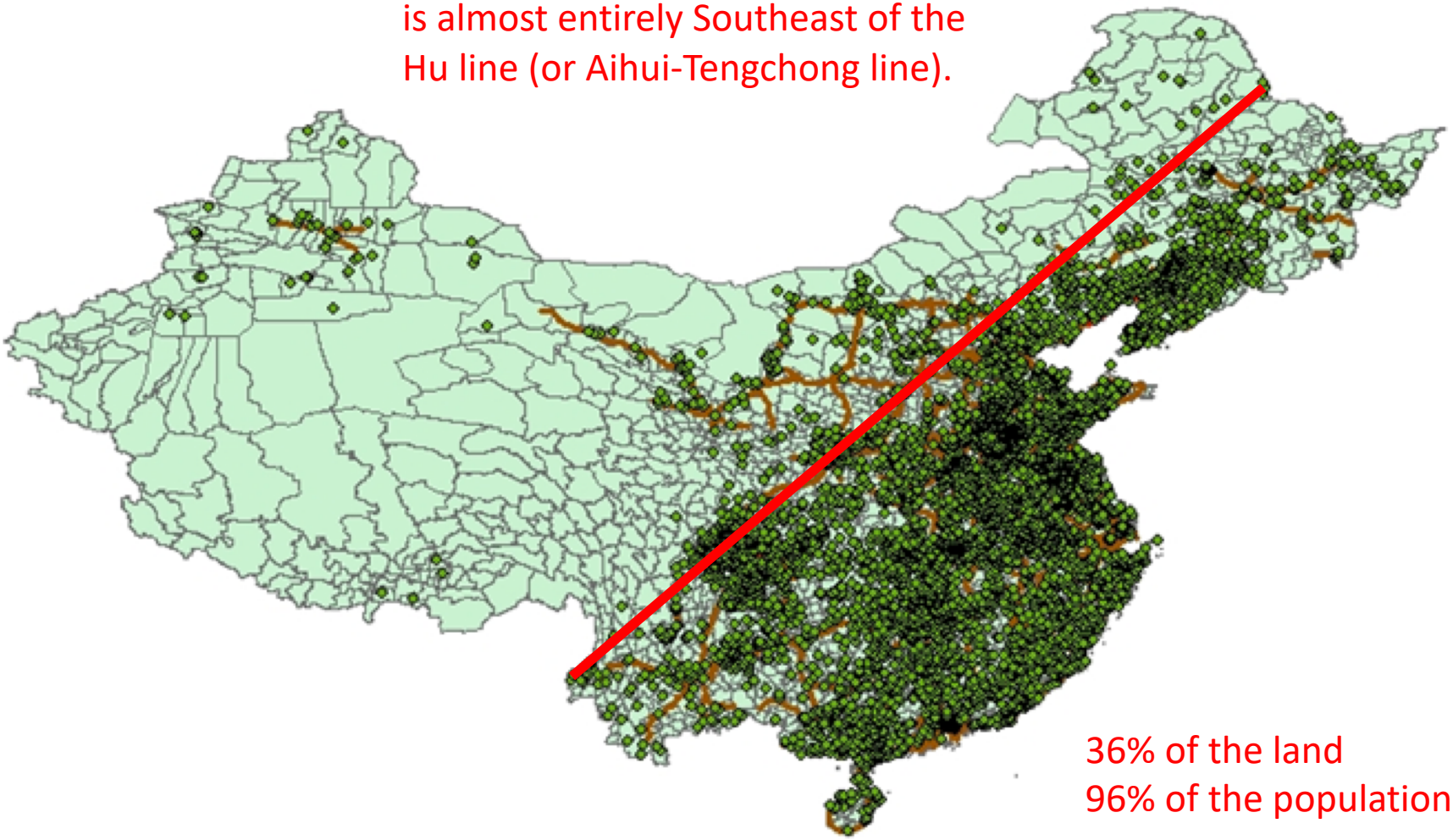
Lower transport costs → easier/cheaper access to inputs (human, material, capital) and easier access to customers. That has direct effects on productivity and intensifies spatial competition. Three productivity effects seem especially relevant in our setting:

- Cheaper inputs (net of transport costs) and lower costs of shipping to customers raises (a) revenue per physical inputs and (b) revenue per net unit input costs
- Better allocation efficiency through selection effects, e.g., exit of least productive plants and firms.
- X-efficiency gains: better coordination across plants of multi-plant firms, sharper managerial incentives to improve plant and firm performance, etc.

Plant locations and national highways as of 1998



Motorways in 2007 and Plants that Entered from 1999 to 2007



Measuring (Revenue) TFP

$\text{Log}TFP_{\text{et}} = \log Q_{\text{et}} - \alpha_{\text{k}}\log K_{\text{et}} - \alpha_{\text{L}}\log L_{\text{et}} - \alpha_{\text{M}}\log M_{\text{et}}$

- *Q*: Total value of shipments plus inventory change (finished products + work-in-progress), deflated by industry-level price index.
- *K:* Net real fixed assets, book value with depreciation
  - Results are similar when we follow Brandt, Van Biesebroeck, and Zhang (2012, 2014) in computing real capital stock.
- *L,* labor input: Total compensation or employment
- *M,* materials: intermediate inputs
- Factor elasticities = industry-level cost shares by year

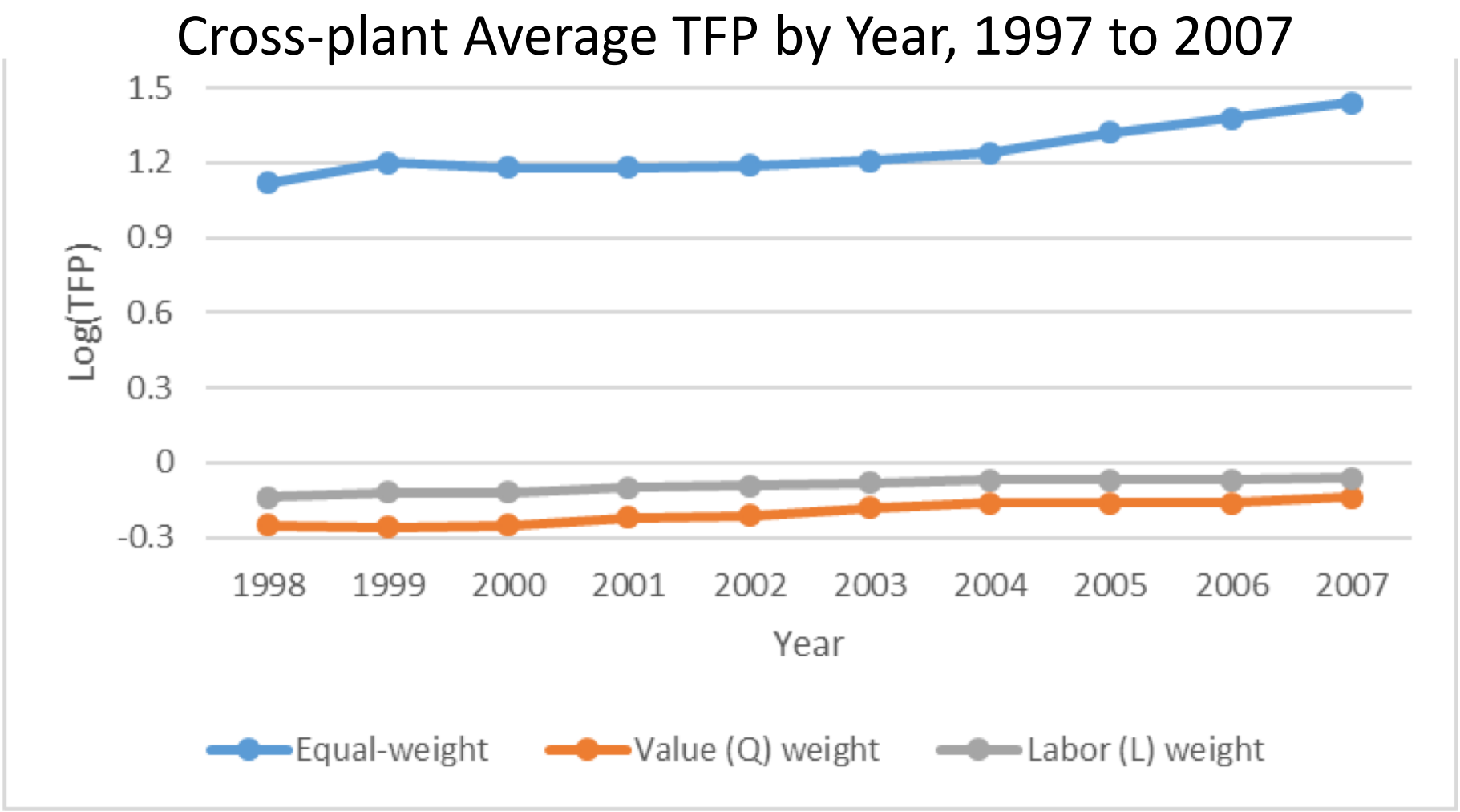
Measuring TFP Deviations

Preliminary step: winsorize TFP at 1st and 99th centiles of TFP by industry-year. Then compute deviations two ways:

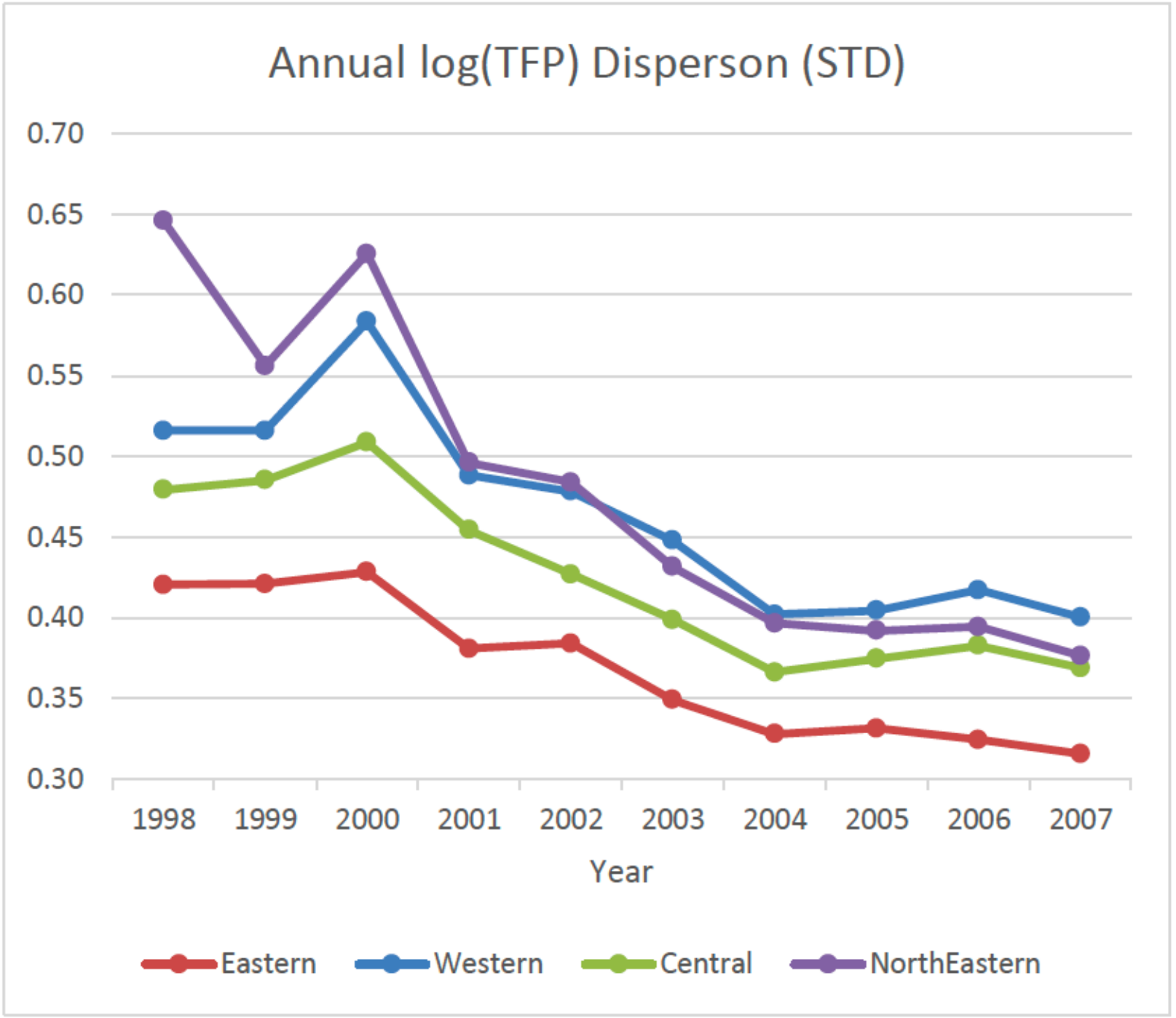
Log Deviations: Log TFP deviations from same-year, same industry mean

Standardized Deviations: First, compute the standard deviation of plant-level log deviations for each industry-year cell. Second, calculate each plant's distance from the industry-year mean log TFP in standard deviation units.

Misallocation in China's Manufacturing Sector



Affirming Hsieh and Klenow (2009): Resources don't flow (enough) to the more productive.


Annual log(TFP) Disperson (STD)
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Year
Eastern
Western
Central
NorthEastern

Standardized Plant-Level TFP Deviation Falls with Distance to Transport Network Access Points

About 2 million plant-year observations in the period from 1997-2007

Dependent Variable: Standardized Plant-level TFP deviation

Key Regressor: Distance from transport mode access point

Controls: Plant size, age, ownership type; province-year FEs; Dummy for >100km from access point

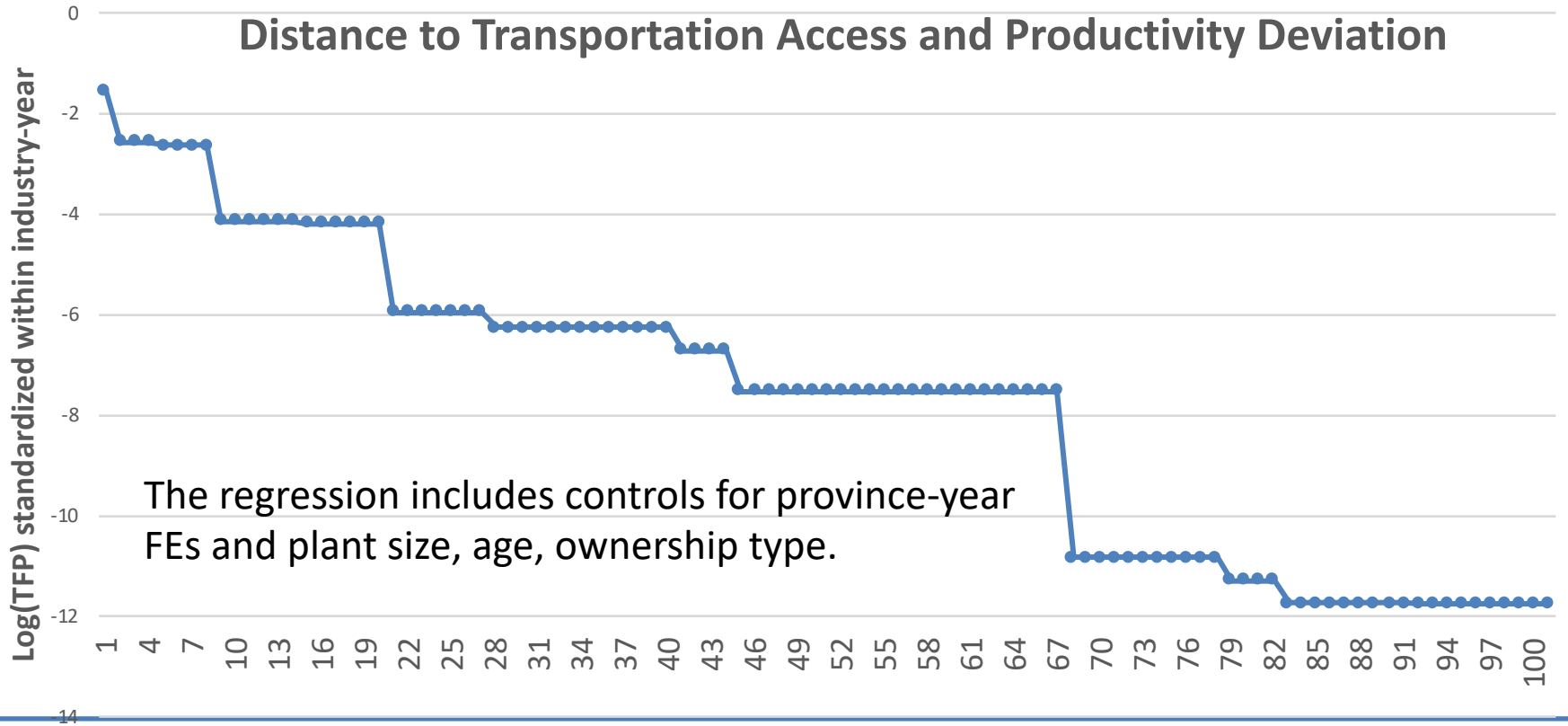
Error clustering at plant-level; similar results when weighting by revenue.

Distance to	Motorway	Railway	Water-way	National highway	Hi-Speed Railway
Y: Standardized deviation, unit in % log(TFP) adjusted by industry mean and standard deviation yearly					
One mode at a time	-2.00***	-12.50***	-4.00***	-17.77***	-1.77
	[-4.46]	[-14.24]	[-8.76]	[-14.84]	[-1.05]
All modes together	-1.80*	-15.50***	-1.99***	-7.35***	0.50***
	[-3.30]	[-16.35]	[-3.75]	[-5.15]	[2.80]

Distance to Motorway Access Point: A Nonparametric Approach

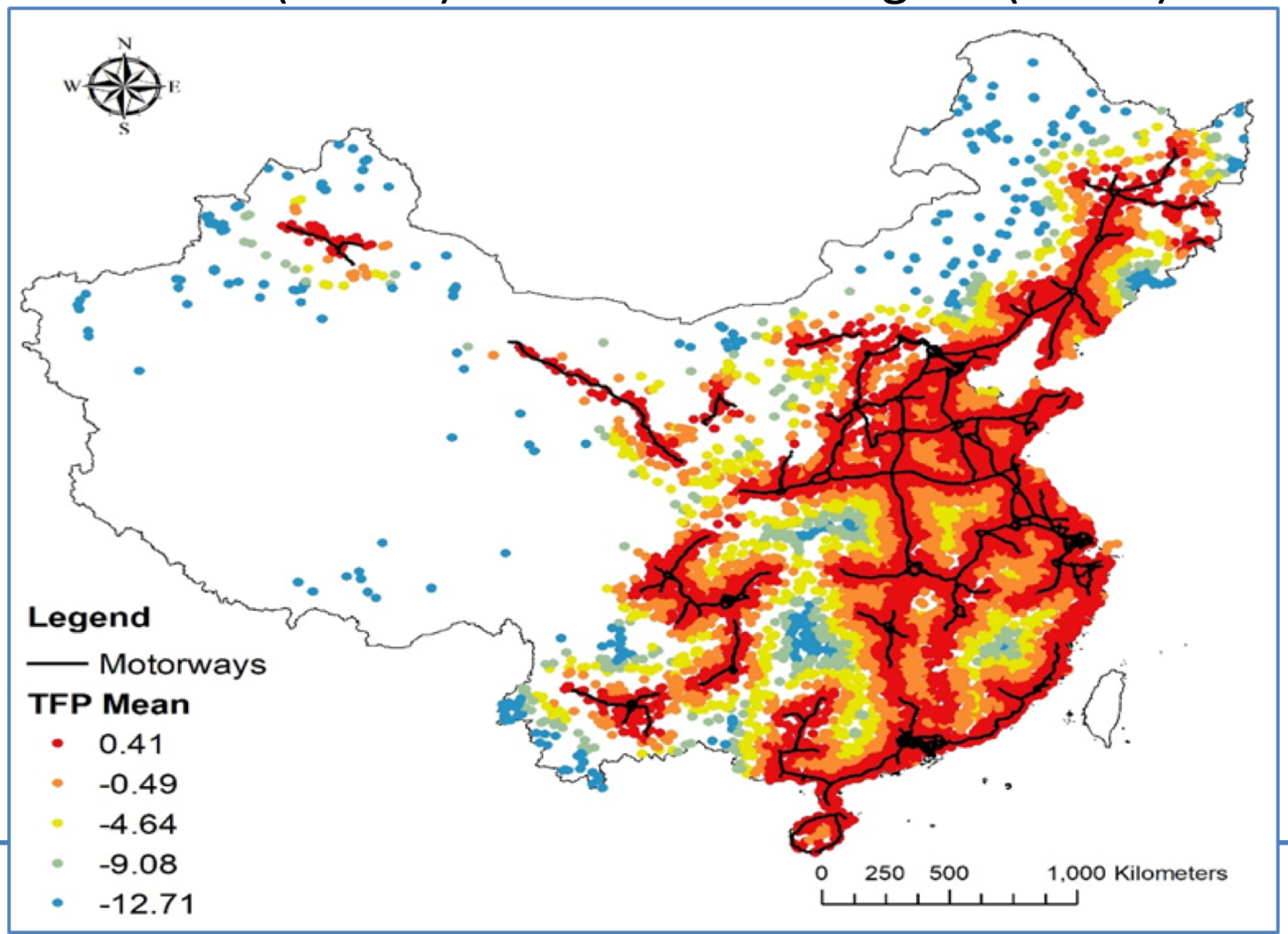
Panel regression with distance bins:

$$\frac{logTFP_{i,s,t} - \overline{logTFP_{s,t}}}{\sigma(LogTFP)_{s,t}} = a + \sum_b \beta_b \, Bin_{i,t,b} + \gamma D(>100km)_{i,t} + +\delta Z + FE_{p,t} + \varepsilon_{i,s,t}$$



TFP Deviation Heat Map

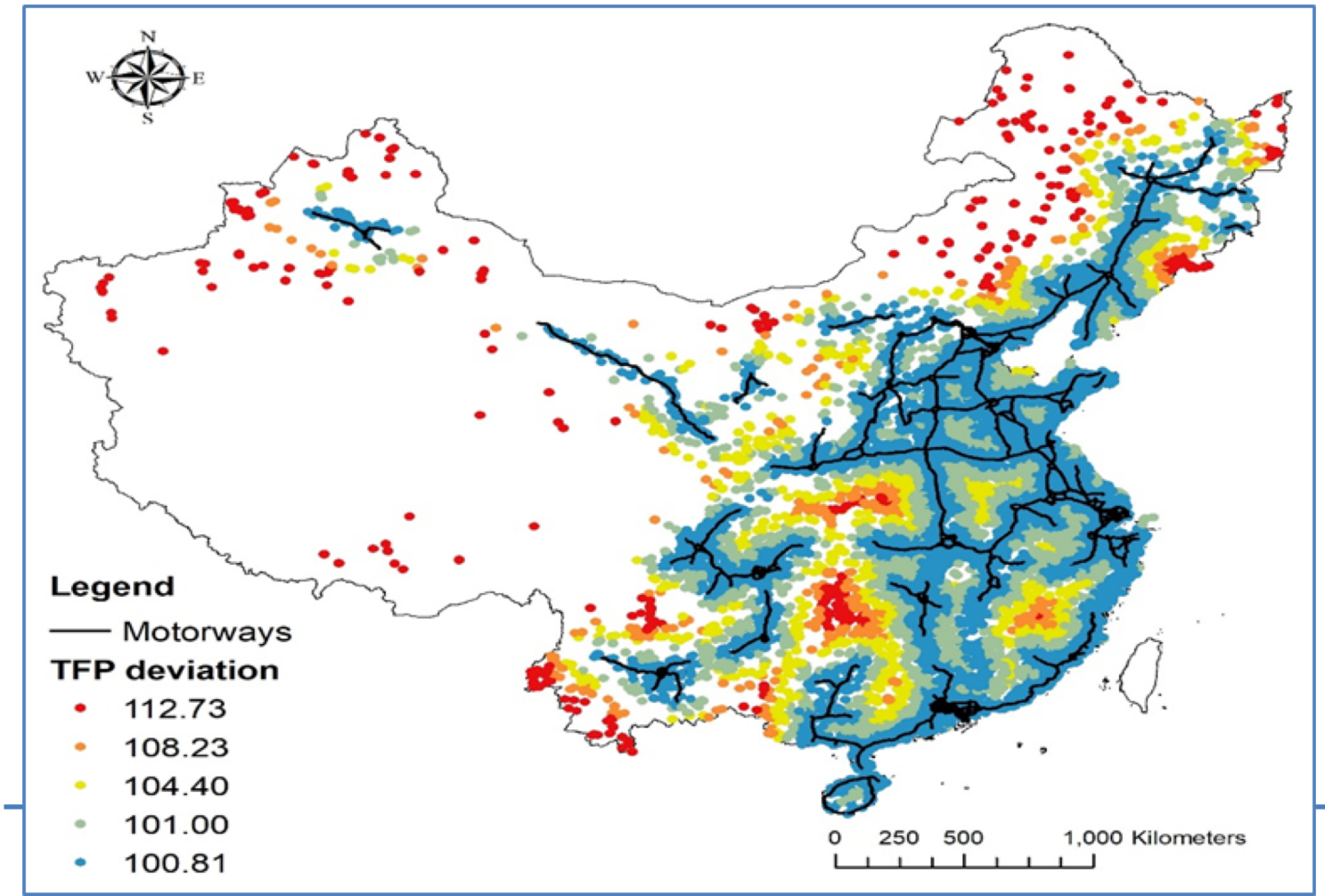
- Plotting fitted values of standardized log(TFP) deviation by zip code from a pooled-sample, plant-level regression on five bins for distance to Motorway access point and same controls as before.
- Warmer (cooler) colour denotes higher (lower) value.



The black line shows the motorway network as of 2004.

TFP Dispersion Heat Map

- Plotting fitted values of abs[standardized log(TFP) deviation] by zip code from a pooled-sample, zip-level regression on five bins for distance to Motorway access point and province-year FEs.
- Warmer (cooler) colour denotes higher (lower) value.



The black line shows the motorway network as of 2004.

First Effort to Quantify Treatment Effects

Treated Units: Plants located within 25km of any new network access point as of year *t* that have no previous network access within 25km.

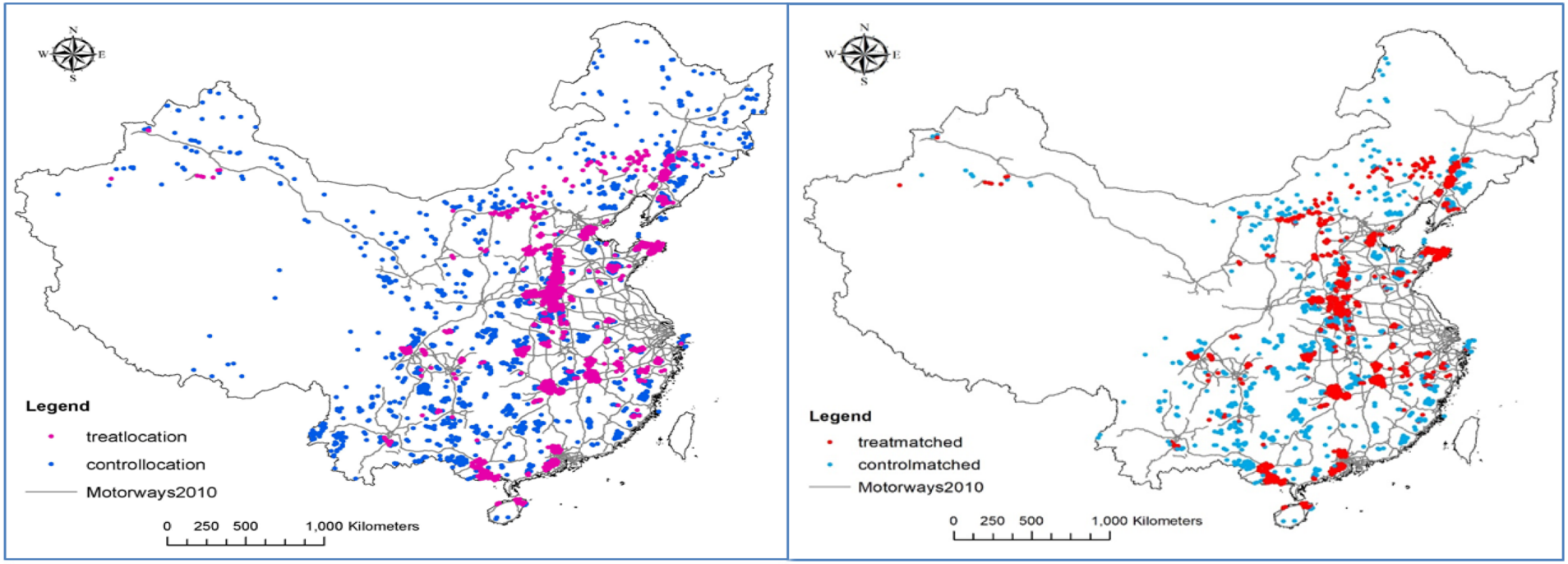
Control Candidates: Plants in the same province and year with no new network access point within 25km from t-2 to t+2 and no previous access within 25km.

Matched Controls: Match each treated unit to the nearest plant among the set of control candidates.

We estimate treatment effects using diff-in-diff specifications.

Repeat using 5km, 10km, and 50km thresholds.

Treated and control locations (zip codes)



We have 38,091 treated units in 1,779 locations. After dropping plants in locations with zero control candidates, we are left with 12,986 treated units in 1,440 locations. The maps above show treated and control locations before (left) and after matching.

For reference, we show the Motorway network in 2010.

Diff-in-Diff Estimated Treatment Effects

Each cell reports results for a separate plant-level regression on the pooled sample, where the dependent variable is the one-year growth rate for the indicated outcome. Little evidence of violation of parallel trends. Treated units grow faster than Controls after treatment. Some evidence, marginally significant, that treated units eventually experience a TFP boost.

From Prior Year to:	Ind-adj log(TFP)	Ind-adj log(TFP)/ σ	Growth Q	Growth M	Growth L	Growth K
t-2	0.29	0.34	-0.09*	0.02	-0.03	0.00
	[1.61]	[1.86]	[1.86]	[0.51]	[0.80]	[0.01]
t-1	-0.11	-0.11	0.01	0.01	-0.07	-0.15**
	[1.10]	[1.06]	[0.22]	[0.16]	[1.56]	[2.10]
T (treat)	0.14	0.12	0.02	0.03	0.12***	0.12**
	[1.23]	[1.05]	[0.73]	[0.86]	[3.71]	[2.31]
t+1	-0.18	-0.21	0.07***	-0.04	-0.03	0.03
	[1.35]	[1.60]	[2.56]	[0.92]	[0.95]	[0.73]
t+2	0.24	0.26	0.01	0.09	-0.03	0.01
	[1.20]	[1.29]	[0.41]	[1.47]	[0.85]	[0.25]
t+3	0.38	0.39	0.01	0.29***	0.05	0.01
	[1.69]	[1.77]	[0.18]	[2.92]	[0.91]	[0.11]

Diff-in-Diff Estimates: 3-Year Changes

	Growth Ind-adj log(TFP)	Growth of Ind-adj log(TFP)/ σ	Growth Q	Growth M	Growth L	Growth K
Changes from t-1 to t+2						
Treatment	-0.27	-0.22	0.09*	0.10**	0.03	0.15*
	[0.69]	[-0.56]	(1.66)	(1.85)	(0.73)	(1.88)
Changes from t to t+3						
Treatment	-1.57	-1.36	0.25***	0.22***	0.18***	0.14
	[-0.27]	[-0.22]	(2.86)	(2.55)	(2.71)	(1.17)

Note: Using DHS growth rates (arc percentage changes)

The two TFP measures are winsorized at 5%

On Plant-Level Entry and Exit

- Faber (2014, RES): Network connections have led to a significant reduction in industrial output growth among non-targeted peripheral counties and falling trade costs between peripheral and metropolitan regions.

➔ Exit rates are higher in the non-treated region, especially those that are skipped in-between highway exits and entrances.

- Brandt et al. (2018, WP), Chen et al. (2018, WP) and Hu et al. (2019, WP): New entries in China are distorted because of policy barriers, selective tax subsidiaries, and corresponding incentives on investment choices.

➔ The improvement in new entries comes from short-term policy rents, not necessarily from better technologies or other sources of efficiency gains.

Treatment & Exit by Location in Own-Industry TFP Distribution

Panel A: Plant exit probabilities in the first two years post transport network access treatment			
Location in own-industry TFP distribution as of the Year t-1	Probability of plant exit by year t+2 (marginal effect)		P-value for difference between the treated and controls
	Treatment	Control	
Bottom Quartile	0.24	0.34	0.16
2nd	0.16	0.21	0.31
3rd	0.18	0.26	0.12
Top	0.17	0.20	0.74

Exit rates fall with position in own-industry TFP distribution. Higher exit rates for Controls might reflect Faber's story. We need to investigate.

Treatment & Entry by Location in Own-Industry TFP Distribution

Panel B: Plants in operation in Year t+3			
Location in Own-Industry TFP Distribution as of the Year t+3	Probability of Plant in operation t+3 enter the market in year t+2 or year t+1		P-value for Difference Between the Treated and Controls
	Treatment	Control	
Bottom Quartile	0.10	0.08	0.07
2nd	0.13	0.11	0.01
3rd	0.14	0.10	0.09
Top	0.13	0.14	0.00

Treatment has only modest effect on entry rates. More work needed.

Summary

1. Tremendous expansion & improvement in China's surface transport network from 1993 to 2020.
2. Enormous dispersion in log (revenue) TFP across Chinese manufacturing plants in same industry.
3. Cross-plant dispersion of log TFP fell sharply from 1998 to 2007.

   -- But the gap between equal-weighted and unweighted average log TFP measures did not shrink, so the lessening of TFP dispersion is not principally on the size dimenion.

Summary, 2

4. Better network access is associated with:
   - Higher plant-level TFP relative to others in same industry-year.
   - Less dispersed plant-level TFP values – again, relative to other plants in the same industry-year.

   Restated: Locations closer to network access have narrower log TFP distributions with higher means.
5. Magnitude: Reducing distance from 80 km to 1 km involves a rise in log TFP of one-tenth of a standard deviation across plants in the same industry-year cell.
6. Initial efforts to estimate treatment effects suggest that building new/closer network access points raises input and output usage for nearby plants, and may raise their TFP. Lots more to do here.

Data Availability

- We are working to make our GIS dataset available to all researchers.
- An open-access dataset of this sort creates a foundation for research in multiple directions and greatly facilitates the study of China's transport network, its expansion over time, and its effects.