"Trade Integration, Market Size, and Industrialization"
Discussion of Faber, ReStud 2014

Nathan Schiff
Shanghai University of Finance and Economics

Graduate Urban Economics, Lecture 6
March 28, 2016

Administration

Homework due today

Main Question

What is the main question?

Main Question

What is the main question?

Main question: what is effect of being connected (new highways) to large cities on small city industrial output?

Main Question

What is the main question?

Main question: what is effect of being connected (new highways) to large cities on small city industrial output?

Two possible effects: 1) production is shifted from large city to connected small city

Main Question

What is the main question?

Main question: what is effect of being connected (new highways) to large cities on small city industrial output?

Two possible effects: 1) production is shifted from large city to connected small city 2) production further concentrates in large city

Main Question

What is the main question?

Main question: what is effect of being connected (new highways) to large cities on small city industrial output?

Two possible effects: 1) production is shifted from large city to connected small city 2) production further concentrates in large city

Additional questions and issues:

1. What happens to population of small city?

Main Question

What is the main question?

Main question: what is effect of being connected (new highways) to large cities on small city industrial output?

Two possible effects: 1) production is shifted from large city to connected small city 2) production further concentrates in large city

Additional questions and issues:

1. What happens to population of small city?
2. How are neighboring, unconnected cities affected?

Innovations and Contributions

Novel “engineering IV”: cleverly used geography as a source of exogenous variation in road placement

Innovations and Contributions

Novel “engineering IV”: cleverly used geography as a source of exogenous variation in road placement

Big question with no theoretical prediction: effect of trade could increase or decrease concentration depending on parameters (Krugman papers)and context (urbanization)

Innovations and Contributions

Novel “engineering IV”: cleverly used geography as a source of exogenous variation in road placement

Big question with no theoretical prediction: effect of trade could increase or decrease concentration depending on parameters (Krugman papers)and context (urbanization)

Quite thorough: results are robust to many specifications, falsification tests, and seem to tell a consistent story

Innovations and Contributions

Novel “engineering IV”: cleverly used geography as a source of exogenous variation in road placement

Big question with no theoretical prediction: effect of trade could increase or decrease concentration depending on parameters (Krugman papers)and context (urbanization)

Quite thorough: results are robust to many specifications, falsification tests, and seem to tell a consistent story

Why study this question in China?

Not entirely clear, my guess:

Why study this question in China?

Not entirely clear, my guess:

China's National Trunk Highway System was built recently and rapidly

Why study this question in China?

Not entirely clear, my guess:

China's National Trunk Highway System was built recently and rapidly

China has many cities and high industrial output, makes identification strategy feasible

Why study this question in China?

Not entirely clear, my guess:

China's National Trunk Highway System was built recently and rapidly

China has many cities and high industrial output, makes identification strategy feasible

Data on city/region output available, geocoded roads, high precision land cover data

Why study this question in China?

Not entirely clear, my guess:

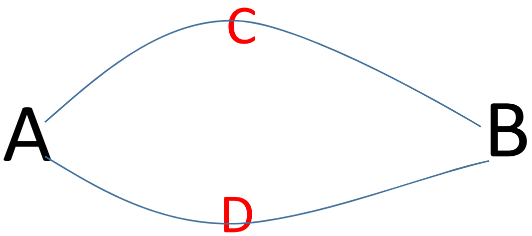
China's National Trunk Highway System was built recently and rapidly

China has many cities and high industrial output, makes identification strategy feasible

Data on city/region output available, geocoded roads, high precision land cover data

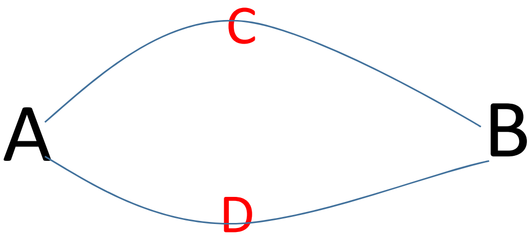
General importance of China as a large country

Basic Setup



New road policy is to connect large cities A and B

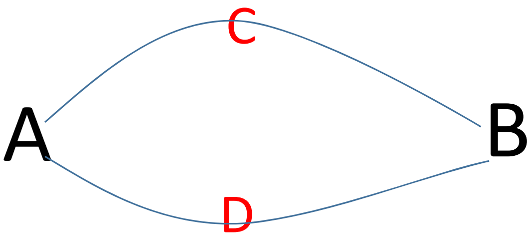
Basic Setup



New road policy is to connect large cities A and B

Road could go through peripheral county (city) C or D

Basic Setup



New road policy is to connect large cities A and B

Road could go through peripheral county (city) C or D

What is effect on peripheral economy of new road?

National Trunk Highway System

Faber writes that policy aim was “to connect all provincial capitals and cities with an urban registered population above 500,000 on a single expressway network, and to construct routes between targeted centres and the border in border provinces as part of the Aisan Highway Network.”

National Trunk Highway System

Faber writes that policy aim was “to connect all provincial capitals and cities with an urban registered population above 500,000 on a single expressway network, and to construct routes between targeted centres and the border in border provinces as part of the Aisan Highway Network.”

- Policy approved in 1992: “7-5” network, 7 horizontal axes, 5 vertical axes

National Trunk Highway System

Faber writes that policy aim was “to connect all provincial capitals and cities with an urban registered population above 500,000 on a single expressway network, and to construct routes between targeted centres and the border in border provinces as part of the Aisan Highway Network.”

- Policy approved in 1992: “7-5” network, 7 horizontal axes, 5 vertical axes
- Constructed between 1992 and 2007 at cost of US$ 120 bn

National Trunk Highway System

Faber writes that policy aim was "to connect all provincial capitals and cities with an urban registered population above 500,000 on a single expressway network, and to construct routes between targeted centres and the border in border provinces as part of the Aisan Highway Network."

- Policy approved in 1992: "7-5" network, 7 horizontal axes, 5 vertical axes
- Constructed between 1992 and 2007 at cost of US$ 120 bn
- 10% open by 1997, additional 81% by 2003,l final 9% after 2003

National Trunk Highway System

Faber writes that policy aim was “to connect all provincial capitals and cities with an urban registered population above 500,000 on a single expressway network, and to construct routes between targeted centres and the border in border provinces as part of the Aisan Highway Network.”

- Policy approved in 1992: “7-5” network, 7 horizontal axes, 5 vertical axes
- Constructed between 1992 and 2007 at cost of US$ 120 bn
- 10% open by 1997, additional 81% by 2003,l final 9% after 2003

Note: no official list of targeted cities, Faber uses stated aim to classify these

National Trunk Highway System map

FIGURE 1

China's National Trunk Highway System. The figure shows Chinese county boundaries in 1999 in combination with the targeted city nodes and the completed expressway routes of the NTHS in the year 2007.

Endogeneity

What is the endogeneity issue here?

Endogeneity

What is the endogeneity issue here?

What is his identification strategy?

Endogeneity

What is the endogeneity issue here?

What is his identification strategy?

Creates two IV road plans:

Endogeneity

What is the endogeneity issue here?

What is his identification strategy?

Creates two IV road plans:

1. Least cost path spanning network: minimize total network cost given cost of building along different land cover types (slope/elevation, developed land, wetland, water

Endogeneity

What is the endogeneity issue here?

What is his identification strategy?

Creates two IV road plans:

1. Least cost path spanning network: minimize total network cost given cost of building along different land cover types (slope/elevation, developed land, wetland, water
2. Euclidean path network: minimize total network cost using straight line connections

Evaluation of IV

What is the relevance requirement and exclusion restriction for this IV?

Evaluation of IV

What is the relevance requirement and exclusion restriction for this IV?

Relevance: must be able to predict placement of road network, *controlling* for other variables (distance to node cities, capital status, urban status, demographics)

Evaluation of IV

What is the relevance requirement and exclusion restriction for this IV?

Relevance: must be able to predict placement of road network, *controlling* for other variables (distance to node cities, capital status, urban status, demographics)

Exclusion restriction: instrument must be uncorrelated with error term; any problems in this context?

Evaluation of IV

What is the relevance requirement and exclusion restriction for this IV?

Relevance: must be able to predict placement of road network, *controlling* for other variables (distance to node cities, capital status, urban status, demographics)

Exclusion restriction: instrument must be uncorrelated with error term; any problems in this context?

Online appendix: discusses testing for whether land cover features could be endogenous (also includes as controls)

Evaluation of IV

What is the relevance requirement and exclusion restriction for this IV?

Relevance: must be able to predict placement of road network, *controlling* for other variables (distance to node cities, capital status, urban status, demographics)

Exclusion restriction: instrument must be uncorrelated with error term; any problems in this context?

Online appendix: discusses testing for whether land cover features could be endogenous (also includes as controls)

Additional discussion of LATE vs population average: argues should be no difference here

Creating Least Cost Path Spanning IV

How does he do this?

Creating Least Cost Path Spanning IV

How does he do this?

- Uses US and Chinese government data on characteristics of land parcels (GIS raster data is basically a grid of cells with values)

Creating Least Cost Path Spanning IV

How does he do this?

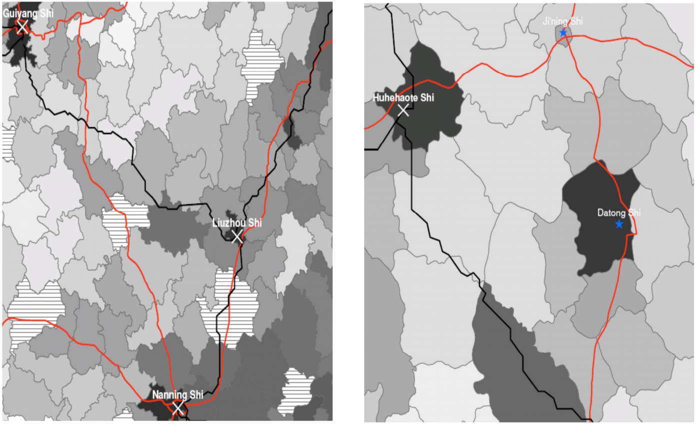
- Uses US and Chinese government data on characteristics of land parcels (GIS raster data is basically a grid of cells with values)
- Uses ArcGIS to run algorithms that minimize construction cost of entire road network based on requirement of connecting 54 cities and given costs of land types

Creating Least Cost Path Spanning IV

How does he do this?

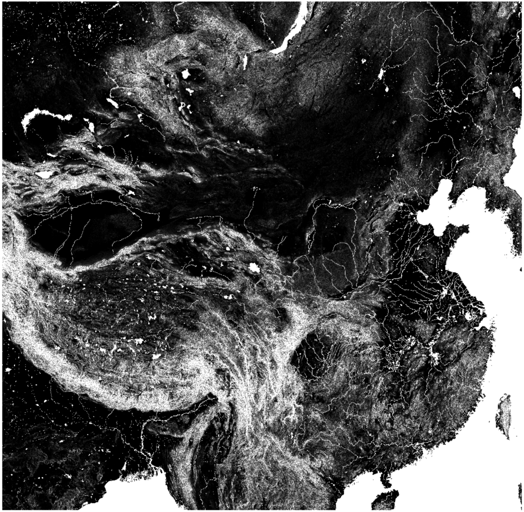
- Uses US and Chinese government data on characteristics of land parcels (GIS raster data is basically a grid of cells with values)
- Uses ArcGIS to run algorithms that minimize construction cost of entire road network based on requirement of connecting 54 cities and given costs of land types
- Output of process is a network of bi-lateral curves between 54 cities

Least Cost Path Spanning IV Example



The network in red color depicts actual NTHS expressway routes. The network in black color depicts the least cost path spanning tree network. Crosses indicate targeted metropolitan nodes. Counties are color coded according to their nominal levels of GDP in 1997, where darker colors represent higher values. Striped areas indicate missing 1997 GDP data.

Construction Cost Map



The figure depicts the construction cost raster used as input into the least cost path algorithm. The color scale ranges from white (very high cost of crossing a parcel of land) to black (very low cost of crossing a square km parcel of land). The cost assignment is based on land gradient (slope) as well as land cover (water, wetlands, and developed land), and described in more detail in the text.

Least Cost Path Spanning IV Map

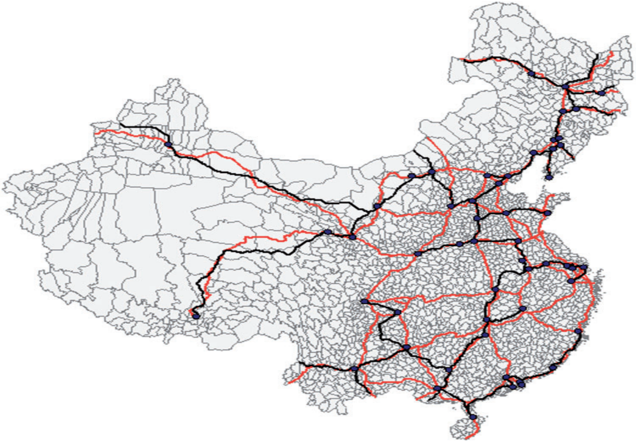


FIGURE 2

Least cost path spanning tree network. The network in red colour depicts the completed NTHS network in 2007. The network in black colour depicts the least cost path spanning tree network. The black routes are the result of a combination of least cost path and minimum spanning tree algorithms. In the first step Dijkstra's (1959) optimal route algorithm is applied to land cover and elevation data in order to construct least costly paths between each bilateral pair of the targeted destination. In the second step, these bilateral cost parameters are fed into Kruskal's (1956) minimum spanning tree algorithm. This algorithm identifies the subset of routes that connect all targeted nodes on a single continuous network subject to global construction cost minimization.

Euclidean IV

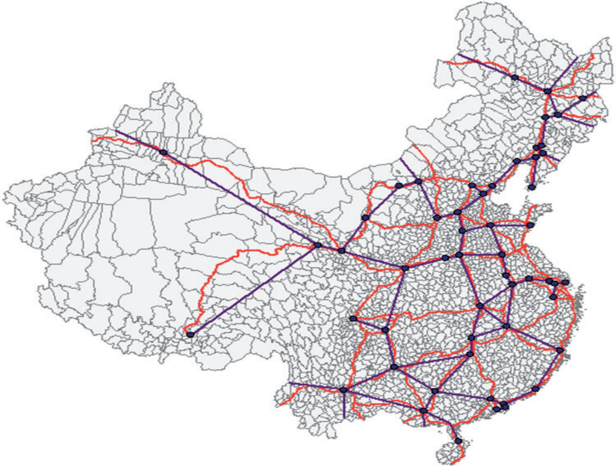


FIGURE 3

Euclidean spanning tree network. The network in red colour depicts the completed NTHS network in 2007. The network in darker colour depicts the Euclidean spanning tree network. The routes are the result of applying Kruskal's (1956) minimum spanning tree algorithm to bilateral Euclidean distances between targeted destinations. This algorithm is first run for the all-China network, and then repeated within North-Centre-South and East-Centre-West divisions of China. These regional repetitions add 9 routes to the original minimum spanning tree.

Main Results

What does he find?

Main Results

What does he find?

Output of connected peripheral counties grows more slowly than non-connected counties

Main Results

What does he find?

Output of connected peripheral counties grows more slowly than non-connected counties

Effects are stronger when including additional controls, suggesting correlation of IV roads and controls

Main Results

What does he find?

Output of connected peripheral counties grows more slowly than non-connected counties

Effects are stronger when including additional controls, suggesting correlation of IV roads and controls

NTHS connections reduced GDP growth by 9% to 18% from 1997-2006; this comes from decrease in industrial output growth (no effect on agriculture)

Main Results

What does he find?

Output of connected peripheral counties grows more slowly than non-connected counties

Effects are stronger when including additional controls, suggesting correlation of IV roads and controls

NTHS connections reduced GDP growth by 9% to 18% from 1997-2006; this comes from decrease in industrial output growth (no effect on agriculture)

No effect on county population growth

Table 2: First Stage

TABLE 2
First stage regressions

Dependent variable:	(1) Connect	(2) Connect	(3) Connect	(4) lnDistHwy	(5) lnDistHwy	(6) lnDistHwy
Least cost path IV	0.323***		0.254***	0.317***		0.245***
	(0.0574)		(0.0635)	(0.0645)		(0.0635)
Euclidean IV		0.243***	0.144**		0.280***	0.193***
		(0.0529)	(0.0560)		(0.0599)	(0.0657)
lnDistNode	−0.130***	−0.127***	−0.104***	0.588***	0.635***	0.426***
	(0.0376)	(0.0295)	(0.0323)	(0.130)	(0.112)	(0.122)
Prefect capital	−0.124*	−0.129*	−0.120*	0.437**	0.429*	0.413*
	(0.0648)	(0.0736)	(0.0658)	(0.209)	(0.229)	(0.215)
City Status	0.0891**	0.0929**	0.0847**	−0.297***	−0.296***	−0.270***
	(0.0403)	(0.0437)	(0.0399)	(0.0946)	(0.103)	(0.0951)
lnUrbPop90	0.106***	0.115***	0.107***	−0.228***	−0.244***	−0.227***
	(0.0225)	(0.0217)	(0.0209)	(0.0691)	(0.0640)	(0.0636)
Educ90	−0.273	−0.303	−0.302	−1.671	−1.747	−1.626
	(0.598)	(0.656)	(0.601)	(1.697)	(1.804)	(1.666)
AgShare90	−0.170	−0.216	−0.167	0.0238	−0.00173	−0.0160
	(0.182)	(0.189)	(0.179)	(0.537)	(0.555)	(0.533)
Obs	1342	1342	1342	1342	1342	1342
R^2	0.222	0.204	0.233	0.401	0.394	0.414
First stage F-Stat	31.61	21.07	20.31	24.09	21.82	15

Notes: All regressions include province fixed effects. Columns 1–3 report results for binary NTHS connection indicators among peripheral counties. Columns 4–6 report results for the log distance to the nearest NTHS segment among peripheral counties. lnDistNode is log county distance to the nearest targeted city node. Prefect Capital and City Status are binary indicators for respective county status in 1990. lnUrbPop90 is log 1990 county urban population. Educ90 is the 1990 county share of above compulsory schooling in 20+ population. AgShare90 is the 1990 county share of agricultural employment. Standard errors are clustered at the province level and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10% significance levels.

Table 3: Main Specification

TABLE 3
Network connection effects among peripheral counties

Dependent variables		(1) OLS No controls	(2) OLS With controls	(3) LCP IV No controls	(4) LCP IV With controls	(5) Euclid IV No controls	(6) Euclid IV With controls	(7) Both IVs No controls	(8) Both IVs With controls	(9) Both IVs With controls
Change ln(IndGVA) 1997–2006	Connect	−0.0529	−0.0356	−0.284**	−0.304**	−0.246*	−0.287*	−0.272***	−0.297***	−0.297**
		(0.0418)	(0.0499)	(0.118)	(0.145)	(0.148)	(0.154)	(0.0965)	(0.108)	(0.121)
	Obs	1302	1280	1302	1280	1302	1280	1302	1280	1280
	R^2	0.242	0.255							
Change ln(NonAgGVA) 1997–2006	Connect	−0.0411	−0.0266	−0.243**	−0.252**	−0.270**	−0.296**	−0.251***	−0.268***	−0.268***
		(0.0335)	(0.0375)	(0.0983)	(0.117)	(0.122)	(0.131)	(0.0877)	(0.0969)	(0.0946)
	Obs	1285	1262	1285	1262	1285	1262	1285	1262	1262
	R^2	0.27	0.284							
Change ln(GovRevenue) 1997–2006	Connect	−0.0497*	−0.0914***	−0.0542	−0.223*	−0.175	−0.315**	−0.0926	−0.257***	−0.257***
		(0.0285)	(0.0295)	(0.109)	(0.120)	(0.117)	(0.132)	(0.0893)	(0.0996)	(0.100)
	Obs	1290	1285	1290	1285	1290	1285	1290	1285	1285
	R^2	0.275	0.334							
Change ln(GDP) 1997–2006	Connect	−0.00204	−0.0144	−0.106	−0.177*	−0.178	−0.254**	−0.127	−0.203**	−0.203**
		(0.0245)	(0.0276)	(0.0830)	(0.0942)	(0.112)	(0.116)	(0.0824)	(0.0886)	(0.080)
	Obs	1297	1272	1297	1272	1297	1272	1297	1272	1272
	R^2	0.228	0.264							
Change ln(AgGVA) 1997–2006	Connect	−0.00344	−0.00790	0.000194	−0.0252	−0.0305	−0.0597	−0.00865	−0.0371	−0.0371
		(0.0210)	(0.0220)	(0.0631)	(0.0789)	(0.0672)	(0.0728)	(0.0545)	(0.0630)	(0.0654)
	Obs	1335	1313	1335	1313	1335	1313	1335	1313	1313
	R^2	0.202	0.208							
Change ln(Population) 1997–2006	Connect	0.00488	−0.00217	0.0395**	0.0264	0.0183	0.0104	0.0333*	0.0207	0.0207
		(0.00456)	(0.00568)	(0.0188)	(0.0234)	(0.0242)	(0.0262)	(0.0183)	(0.0215)	(0.0225)
	Obs	1337	1314	1337	1314	1337	1314	1337	1314	1314
	R^2	0.234	0.271							

Table 4: Falsification Test

TABLE 4
Falsification test before and after the network was built

Dependent variable: Change ln(LocGovRev)	(1) OLS 1990–97	(2) OLS 1997–06	(3) LCP IV 1990–97	(4) LCP IV 1997–06	(5) Euclid IV 1990–97	(6) Euclid IV 1997–06	(7) Both IVs 1990–97	(8) Both IVs 1997–06
Panel A: Binary								
Connect	0.0154	−0.0848**	0.0143	−0.151	0.117	−0.282**	0.0563	−0.204***
	(0.0410)	(0.0360)	(0.0853)	(0.0974)	(0.107)	(0.129)	(0.0647)	(0.0467)
Obs	894	894	894	894	894	894	894	894
R^2	0.274	0.339						
First stage F-Stat			19.635	19.635	19.091	19.091	14.93	14.93
Panel B: log Distance								
ln(DistHwy)	−0.0114	0.0160	−0.0409	0.0854*	−0.00442	0.185**	−0.0274	0.122***
	(0.0142)	(0.0190)	(0.0350)	(0.0470)	(0.0573)	(0.0783)	(0.0329)	(0.0430)
Obs	894	894	894	894	894	894	894	894
R^2	0.275	0.336						
First stage F-Stat			18.696	18.696	17.306	17.306	11.259	11.259

Notes: Each point estimate stems from a separate regression. All regressions include province fixed effects and a full set of county controls. LCP IV stands for the least cost path spanning tree instrument. Euclid IV stands for the straight line spanning tree instrument. Panel A presents results for binary NTHS connection indicators (for both OLS and instruments) and Panel B presents results for log distance to the nearest NTHS segment (again for both OLS and instruments). Standard errors are clustered at the province level and stated in parentheses below point estimates. ***1%, **5%, and *10% significance levels.

Two Possible Explanations

What are the two explanations for his results that he tests?

Two Possible Explanations

What are the two explanations for his results that he tests?

1) Trade effect: bigger market with IRS production leads to concentration of production in core, periphery loses (home market effect)

Two Possible Explanations

What are the two explanations for his results that he tests?

1) Trade effect: bigger market with IRS production leads to concentration of production in core, periphery loses (home market effect)

This effect can be reinforced when workers move to core area (last class)

Two Possible Explanations

What are the two explanations for his results that he tests?

1) Trade effect: bigger market with IRS production leads to concentration of production in core, periphery loses (home market effect)

This effect can be reinforced when workers move to core area (last class)

2) Urbanization effect: connected counties lose industry to nearby unconnected counties (decentralization)

Two Possible Explanations

What are the two explanations for his results that he tests?

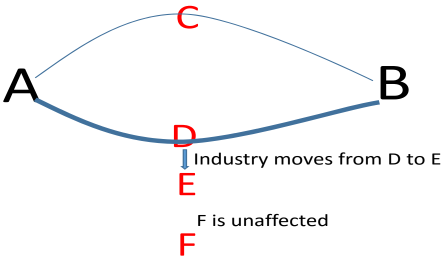
1) Trade effect: bigger market with IRS production leads to concentration of production in core, periphery loses (home market effect)

This effect can be reinforced when workers move to core area (last class)

2) Urbanization effect: connected counties lose industry to nearby unconnected counties (decentralization)

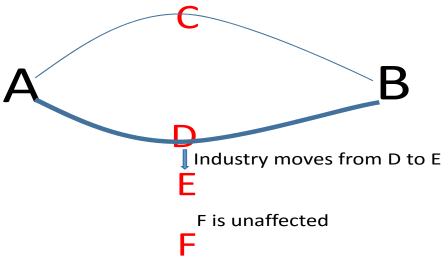
Also implies population growth should differ in connected counties compared to nearby unconnected (but how exactly? not a good fit for monocentric city model)

Effect of Distance to NTHS Road



Decentralization implies non-monotonic effect of distance to road on output

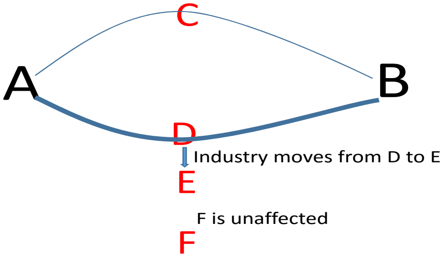
Effect of Distance to NTHS Road



Decentralization implies non-monotonic effect of distance to road on output

D should be negatively affected, E positive, F unchanged

Effect of Distance to NTHS Road



Decentralization implies non-monotonic effect of distance to road on output

D should be negatively affected, E positive, F unchanged

To test this estimates effect of distance using high-order polynomial (effect can vary greatly with distance)

Fitted Effect of Distance

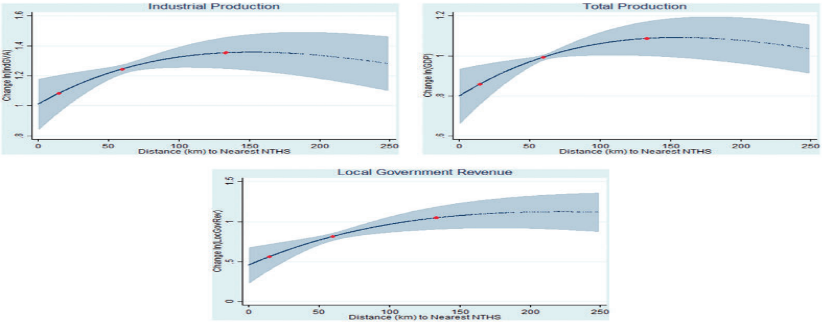


FIGURE 4

Estimated effect of peripheral connections over distance to the nearest NTHS route. The graphs depict the flexibly estimated relationships between distance to the nearest NTHS route and peripheral county growth in industrial value added, total GDP, and local government revenue. The plots correspond to the best fitting polynomial functional form according to the Akaike Information Criterion (AIC). The functions and confidence intervals are based on IV estimates holding covariates at their mean. County distance to the NTHS and its polynomial terms are instrumented with distances to the LCP and Euclidean spanning trees and their polynomials. The red dots indicate median county distances to the nearest NTHS route among connected peripheral counties (left), peripheral counties neighbouring a connected county (centre), and the remaining peripheral counties farther away (right). The shaded areas indicate 90% confidence intervals. Standard errors are clustered at the province level.

Discussion

Do you find results convincing? Surprising?

Discussion

Do you find results convincing? Surprising?

Are you satisfied with the instruments?

Discussion

Do you find results convincing? Surprising?

Are you satisfied with the instruments?

Anything you would do differently?

Discussion

Do you find results convincing? Surprising?

Are you satisfied with the instruments?

Anything you would do differently?

Any important details of Chinese context ignored?

Discussion

Do you find results convincing? Surprising?

Are you satisfied with the instruments?

Anything you would do differently?

Any important details of Chinese context ignored?