Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System*

Benjamin Faber†

February 21, 2014

Abstract

Large scale transport infrastructure investments connect both large metropolitan centers of production as well as small peripheral regions. Are the resulting trade cost reductions a force for the diffusion of industrial and total economic activity to peripheral regions, or do they reinforce the concentration of production in space? This paper exploits China's National Trunk Highway System as a large scale natural experiment to contribute to our understanding of this question. The network was designed to connect provincial capitals and cities with an urban population above 500,000. As a side effect, a large number of small peripheral counties were connected to large metropolitan agglomerations. To address non-random route placements on the way between targeted city nodes, I propose an instrumental variable strategy based on the construction of least cost path spanning tree networks. The estimation results suggest that network connections have led to a reduction in GDP growth among non-targeted peripheral counties. This effect appears to be driven by a significant reduction in industrial output growth. Additional results present evidence in support of a trade based channel in the light of falling trade costs between peripheral and metropolitan regions.

Keywords: Market integration; transport infrastructure; core-periphery; industrialization
JEL Classification: F12; F15; O18; R12

*I am grateful to Robin Burgess, Dave Donaldson, Henry Overman, Steve Redding and Daniel Sturm for their advice and encouragement, and to Lawrence Crissman at the ACASIAN Data Center at Griffith University for his support with the GIS data. The paper benefited from conversations with Esther Duflo, Alex Lembcke, Thierry Mayer, Guy Michaels, Nancy Qian, Steve Pischke, Frederic Robert-Nicoud, Thomas Sampson and Matti Sarvimaki, seminar audiences at the ASSA, IGC, LSE, MIT, UC Irvine and Warwick, and the comments of four anonymous referees and the Editor. The research was supported by the UK Economic and Social Research Council.

†Department of Economics, University of California Berkeley; Email: benfaber@econ.berkeley.edu.

1 Introduction

A large share of world trade takes place between regions within countries. In this context, transport infrastructure investments have been a prominent policy tool that directly affects the degree of within-country trade integration. These policies frequently combine national efficiency with regional equity objectives under the presumption that falling trade costs promote both national growth as well as the diffusion of economic activity to peripheral regions.[1]

Large scale transport investments almost inevitably connect both large metropolitan centers of production as well as small peripheral regions. This is especially the case in developing countries where spatial disparities are particularly pronounced (Kanbur and Venables, 2005). Are the resulting trade cost reductions between *ex ante* asymmetric markets a force for the diffusion of industrial and total economic activity to peripheral regions, or do they reinforce the concentration of production in space? Despite widespread policy interest in this question, our existing empirical knowledge is limited. The growing body of empirical literature on the evaluation of transport infrastructure has so far paid little attention to the role of asymmetric market sizes in trade integration, and the policy question of how peripheral regions are affected when connected to large metropolitan agglomerations.

This paper studies China's National Trunk Highway System (NTHS) as a large scale natural experiment to contribute to our understanding of this question. The NTHS policy objective was to connect provincial capitals, cities with an urban population above 500,000, and border crossings on a single expressway network (World Bank, 2007b). While the targeted metropolitan centers represented just 1.5% of China's land area and 17% of its population, they accounted for 60% of the country's non-agricultural production before the bulk of the network was built in 1997. The average pre-existing difference in market sizes between non-targeted peripheral counties and targeted metropolitan centers was 1:25 in terms of county GDP. The NTHS thus provides an empirical setting where, as a by-product of the network's policy objective, a large number of relatively small peripheral counties were connected to China's major centers of production.

The analysis exploits this policy setting to empirically estimate the economic consequences of NTHS network connections among peripheral counties on the way between targeted metropolitan centers. While peripheral counties were not explicitly targeted by the policy, it would nevertheless be a strong *ex ante* assumption that route placements on the way between targeted city centers were randomly assigned. In particular, both the NTHS planning process and descriptive statistics suggest that planners favored politically important and economically prosperous peripheral counties on the way between targeted nodes. To address these concerns I propose an instrumental variable (IV) strategy based on hypothetical least cost path spanning tree networks. These correspond to the question which routes planners would have been likely to build if the sole policy objective had been to connect all targeted city nodes on a single continuous network subject to global construction cost minimization. In a first step, I use remote sensing data on land cover and elevation in combination with Dijkstra's (1959) optimal route algorithm to compute least costly

[1] For example, from the World Bank's Transport Business Strategy 2008-2012 (2008, pp. 03): "*One of the best ways to promote rural development is to ensure good accessibility to growing and competitive urban markets.*" For discussions of the bundling of efficiency with regional equity objectives see for example Puga (2002) and Ottaviano (2008).

construction paths between any bilateral pair of targeted nodes. I then use these bilateral cost parameters in combination with Kruskal's (1956) minimum spanning tree algorithm to identify the subset of routes that connect all targeted nodes on a single continuous graph subject to network construction cost minimization. I also construct a more extensive, but less precise Euclidean straight line spanning tree network that is subject to a different trade-off between route precision and the number of captured bilateral connections.

The baseline identifying assumption is that county location along an all-China least cost spanning tree network affects changes of county level economic outcomes only through NTHS highway connections, conditional on province fixed effects, distance to the nearest targeted city node, and controls for pre-existing political and economic characteristics. To assess the validity of this assumption, I report how baseline IV point estimates are affected by the inclusion of county controls, and test the heterogeneity of the connection effects across pre-existing county characteristics. In addition, the relatively recent nature of the NTHS allows me to test for network connection effects on identical county samples both before and after the network was built as a placebo falsification test.

The estimation results suggest that NTHS connections have led to a reduction in local GDP growth among peripheral counties on the way between targeted metropolitan centers relative to non-connected peripheral counties. This effect appears to be driven by a significant reduction in industrial output growth. These results are confirmed on local government revenue growth instead of production data. Furthermore, network connections do not appear to have significant effects on county populations.

After documenting the reduced form connection effects, the analysis proceeds to further investigate the channels at work. The main question that I address is whether the observed effects are driven by a mechanism based on inter-regional trade cost reductions between connected peripheral and metropolitan regions, as opposed to a mechanism based on the local effects of the NTHS between connected and neighboring non-connected peripheral counties. On one side, the observed effects could be driven by inter-regional trade integration. In particular, the results could be accounted for by core-periphery effects of trade integration found in increasing returns to scale trade theory and economic geography (Helpman and Krugman, 1985; Fujita *et al.*, 1999). A trade based channel could also be consistent with neoclassical locational fundamentals/comparative advantages between peripheral and metropolitan regions in the presence of barriers to the reallocation of factors (Goldberg and Pavcnik, 2007; Topalova, 2010) or capital mobility (Redding, 2012). On the other side, a mechanism based on the local effects of transport infrastructure could be motivated by urban economics theory (Alonso, 1964; Muth, 1969) and empirical findings in Baum-Snow (2007) and Baum-Snow *et al.* (2012). In particular, instead of inter-regional trade cost reductions, NTHS placements could be associated to a local process of urbanization and industrial decentralization from connected to neighboring non-connected peripheral counties.

Motivated by the testable implications of the different mechanisms, I present a number of additional estimation results. The main findings are that: i) NTHS connections do not appear to have significant effects on changes in urban population or urbanization among connected peripheral counties relative to non-connected ones; ii) The estimated adverse effects on production and government revenue growth do not appear to be driven by a process of local decentralization of production to surrounding areas in the neighborhood of connected peripheral counties; and iii) Peripheral counties with larger pre-existing market sizes and higher pre-existing trade costs *vis-à-vis* the metropolitan nodes are significantly less adversely affected by NTHS connections.

The additional results appear to be consistent with a trade based channel in the light of falling trade costs between *ex ante* asymmetric regions, and suggest that large scale inter-regional transport infrastructure can lead to a different set of economic consequences than those observed when analyzing the local effects of transport projects within metropolitan areas. These findings relate to existing theoretically motivated policy discussions in the trade literature (Baldwin *et al.*, 2003; Combes *et al.*, 2008; Ottaviano, 2008), and serve to emphasize the importance of potentially unintended general equilibrium consequences when evaluating and planning large scale transport policies.

This research is related to a growing empirical literature on the evaluation of transport infrastructure. Recent contributions have studied the economic effects on suburbanization (Baum-Snow, 2007), skill premia in local labor markets (Michaels, 2008), long term GDP effects (Banerjee *et al.*, 2012), gains from trade (Donaldson, 2013), urban form (Baum-Snow *et al.*, 2012), and city growth (Duranton and Turner, 2012). Relative to the existing literature, this paper draws attention to a different question of policy interest. In particular, I study China's NTHS policy as a large scale natural experiment to learn about the role of asymmetric market sizes in trade integration, and to provide empirical evidence on the question of how relatively small peripheral market places are affected by transport connections to large metropolitan centers of production.

The empirical strategy of this paper is related to recent empirical work on transport infrastructure in China (Banerjee *et al.*, 2012; Baum-Snow *et al.*, 2012). It is most closely related to Banerjee *et al.* (2012) who use straight line connections between nearest-neighbor pairs of Treaty Ports and historical cities in China to predict the construction of railway lines in the late 19th and early 20th century. Relative to existing work, this paper exploits the network design of the NTHS policy and its targeting of a subset of major city nodes to propose an IV strategy that is based on minimum spanning tree networks. In addition to instrumenting for route placements on any given bilateral connection, such as by straight lines in Banerjee *et al.* (2012) or by applying Dijkstra's algorithm on land cover and elevation data as in this paper, the spanning tree approach allows to instrument for the choice of bilateral route connections covered by the network subject to global construction cost minimization. This strategy can provide a useful empirical tool in the evaluation of a variety of different infrastructure policies whose design is based on targeting subsets of centers on national or regional grids, such as transportation networks, utility grids, electricity grids or telecommunication networks.

The remainder of the paper is structured as follows. Section 2 describes the policy background and data. Section 3 presents the empirical strategy. Section 4 presents estimation results. Section 5 presents additional estimation results to further investigate the channels at work. Section 6 concludes.

2 Background and Data

2.1 China's National Trunk Highway System

In 1992, the Chinese State Council approved the construction of the "7-5" network, consisting of seven horizontal and five vertical axes, under the National Trunk Highway Development Program (Asian Infrastructure Monthly, 1995; World Bank, 2007b) (see Figure 1). The NTHS was constructed at an estimated cost of US$ 120 billion over a 15-year period until the end of 2007, spanning approximately 35,000 km of high speed four-lane highways (Li and Shum, 2001; Asian Development Bank, 2007; World Bank, 2007a).

Its stated objectives were 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 centers and the border in border provinces as part of the Asian Highway Network. NTHS routes are four-lane limited access toll ways. Road quality, congestion, and driving speed of the modern expressways are in clear contrast to pre-existing national highways and provincial highways that can also be subject to road tolls.[2]

The network was originally earmarked for completion by 2020, but was completed ahead of schedule by the end of 2007. Planners at the Chinese Ministry of Communications divide the construction into a "kick-off" phase between 1992-1997, and "rapid development" between 1998-2007 (World Bank, 2007a). The reason behind the acceleration of construction efforts in 1998 is that highway construction became part of the government's stimulus spending after the Asian financial crisis (Asian Development Bank, 2007).

To finance the great majority of NTHS routes, the central government encouraged province and county level governments to raise funds by borrowing against future toll revenues.[3] Construction was undertaken almost entirely by Chinese state owned enterprises, part of which were assigned directly to particular localities, part of which were participating in contract auctions.[4] Given the progress in the construction of the NTHS ahead of plan, the State Council approved an even more ambitious follow-up blue print for highway construction in 2004. The so called "7-9-18" system has the stated objective to connect all cities with an urban registered population of more than 200,000. It is scheduled to be completed by 2020.

2.2 Data

This section describes the data and variables used in the estimations. A more detailed description can be found in the Online Appendix. Geo-referenced administrative boundary data for the year 1999 were obtained from the ACASIAN Data Center at Griffith University in Brisbane, Australia. These data provide a county-level geographical information system (GIS) dividing the surface of China into 2341 county level administrative units, 349 prefectures, and 33 provinces. Chinese

[2]The common NTHS speed limit is 120 km/h, and a common minimum speed limit is 70km/h. Pre-existing national and provincial highways are typically subject to 80-100km/h and 70km/h respectively.

[3]Roughly 70% was financed from province and county level debt, and 10-15% was contributed by the central government. Private sector participation was also encouraged with up to 5% of financing stemming from domestic and foreign investors (Asian Development Bank, 2007).

[4]Until the "Measures on Tenders and Bids for Contracts for Construction Projects" came into effect in May 2003, competitive bidding, was recommended but not mandatory (World Bank, 2007a).

administrative units at the county level are subdivided into county level cities (shi), counties (xian), and urban wards of prefecture level cities (shixiaqu).[5]

County level socioeconomic records are taken from Provincial Statistical Yearbooks for the years 1990, 1997 and 2006, as well as the 1990 Chinese population census. The statistical yearbook records for 1997 and 2006 were obtained from the University of Michigan's China Data Center, and the 1990 census data as well as statistical yearbook data for 1990 were obtained from the China in Time and Space (CITAS) project at the University of Washington. The Provincial Statistical Yearbook series report county level GDP broken up into agriculture, industry, and services gross value added, as well as local government revenues and registered county populations. The 1990 Population Census provides county level data on population, education, and employment shares by sector.

These sources result in a database of 1748 historically consistent geo-referenced county units that have non-missing reporting values in the Provincial Statistical Yearbooks of 1997 and 2006 (75% of Chinese administrative units). Close to the entirety of this county sample (1706 of 1748) also report socioeconomic records in the 1990 Population Census, and 1238 of the 1748 report local government revenues in the CITAS Provincial Statistical Yearbooks for the year 1990.[6] Table 1 presents a set of descriptive statistics.

Geo-referenced NTHS highway routes were obtained from the ACASIAN Data Center. NTHS routes were digitized on the basis of a collection of high resolution road atlas sources published between 1998 and 2007. These atlas sources made it possible to classify NTHS segments into three categories that coincide with the main construction phases described by the Ministry of Communications: opened to traffic before mid-1997 (10% of NTHS), opened to traffic between mid-1997 and end of 2003 (81% of NTHS), and opened to traffic after 2003 (9% of NTHS). Finally, land cover and elevation data that are used in the construction of least cost path highway routes were obtained from the US Geological Survey Digital Chart of the World project, and complemented by higher resolution Chinese hydrology data from the ACASIAN data center.

3 Empirical Strategy

The data described in the previous section is used to estimate the effects of NTHS network connections among peripheral counties between 1992-2003 on changes of economic outcomes between 1997-2006. The baseline estimation strategy is a difference in differences specification of the form:

$$\ln\left(y_{ip}^{2006}\right)\text{-}\ln\left(y_{ip}^{1997}\right)=\gamma_p+\ \beta Connect_{ip}+\eta \mathrm{X}_{ip}+\epsilon_{ip} \tag{1}$$

where y_{ip} is an outcome of interest of county i in province p, γ_{p} is a province fixed effect, $Connect_{\mathrm{ip}}$ indicates whether i was connected to the NTHS between 1992-2003, and X_{ip} is a vector of county control variables described below. Because targeted city regions encompass multiple county level units, I exclude county observations within a 50 km commuting radius around the targeted city

[5] In case of multiple central city wards (shixiaqu) of a prefecture level city, these are reported as one county level administrative unit.

[6] Only a fraction of the reporting counties in 1997 and 2006 report production data in the Provincial Statistical Yearbooks for 1990.

centers.[7] I classify highway connections using GIS with a dummy indicator that takes the value of one if any part of county i is within a 10km distance of a NTHS highway that was opened to traffic before the end of 2003. Alternatively, I run specification (1) with a continuous treatment variable, $\ln DistHwy_{\text{ip}}$, which stands for the logarithm of great circle distance to the nearest NTHS highway segment opened to traffic before the end of 2003, measured from the center of each county unit. Given that 89% of the reported NTHS connections until the end of 2003 were completed during the phase of "rapid development" between 1998-2003, the main source of variation used in the estimations stems from network connections during this five year period and their effects on changes of economic outcomes over the nine year period 1997-06. The error term ϵ_{ip} could be correlated across counties that were connected to a similar part of the network during a similar period between 1992-2003. I therefore cluster standard errors at the level of 33 Chinese provinces. Alternatively, I follow Conley (1999) and allow for spatial dependence to be a declining function over bilateral county distances without imposing parametric assumptions.

3.1 Least Cost Path Spanning Tree Networks

Estimating specification (1) by OLS would imply the assumption that county connections between nodal cities were randomly assigned within provinces. Given the policy setting of the NTHS, this assumption would be strong. The NTHS was planned in 1992 to establish the backbone of a modernized road transport system for China. Province and county governments borrowed against future toll revenues to finance its construction. This background raises the concern that planners targeted politically important and economically prosperous regions on the way between the network's nodal cities. This concern is supported by descriptive statistics reported in Table 1. Peripheral counties connected to the network by 2006 were on average larger, richer, more urbanized, and more industrialized than non-connected peripheral counties in 1997 before the bulk of the network had been built.[8]

To address these concerns, I construct two hypothetical minimum spanning tree highway networks as instruments for actual route placements (see Figures 2 and 3). I refer to the first as least cost path spanning tree network, and to the second as Euclidean spanning tree network. Both instruments correspond to the question of which routes central planners would have been likely to construct if the sole policy objective had been to connect all targeted destinations on a single network subject to global construction cost minimization. As I outline below, the least cost path network yields more precise route predictions between any given bilateral connection on an all-China minimum spanning tree due to its use of land cover and elevation data, while the Euclidean network covers a larger set of the actually built bilateral network routes.

The following paragraphs describe the spanning tree instruments, and a more detailed description is provided in the Online Appendix. The least cost path network depicted in Figure 2 is constructed in a two-step procedure. The first step is to compute least cost highway construction paths between all possible targeted destination pairs on the basis of remote sensing data on land cover and elevation. To this end, I adopt a simple construction cost function from the transport engineering literature that assigns higher construction costs to land parcels with steeper slope gra-

[7] See Garske *et al.* (2011) for a study of commuting patterns in China.
[8] Reported differences are statistically significant at the 1% level.

dients and land cover classified as water, wetlands, or built structures (Jha *et al.*, 2001; Jong and Schonfeld, 2003).[9] I use the remote sensing data to create a construction cost surface covering the PR China in a rectangular grid of cost pixels.[10]

I then implement Dijkstra's optimal route algorithm to construct least cost highway construction paths between all possible bilateral city connections as well as provincial capitals and the border in border provinces. In the second step, I extract the estimated construction cost of each bilateral connection and feed them into Kruskal's minimum spanning tree algorithm. This algorithm identifies the subset of routes that connect all targeted cities on a single continuous network subject to global construction cost minimization. The resulting number of bilateral connections covered by the minimum spanning tree is equal to the number of nodes minus 1.

To construct the Euclidean spanning tree network depicted in Figure 3, the first step is to compute great circle distances between all possible bilateral connections of the network. I then compute Kruskal's algorithm to identify the minimum number of edges that connect all targeted destinations subject to the minimization of total network distance. To compensate for the loss of route precision, I account for the fact that Chinese planners constructed many more than the minimum number of spanning tree connections. I therefore re-run Kruskal's algorithm within separate geographical subdivisions after dividing China into North-Center-South or East-Center-West geographical divisions.[11] These six additional spanning tree computations add nine bilateral routes in addition to the single all-China minimum spanning tree.

3.2 Additional County Controls and Identifying Assumption

The minimization of a nationwide network construction cost objective function from which the instruments in Figures 2 and 3 are derived is aimed to address the concern of non-random local route placements on the way between targeted city nodes. However, the exogeneity of the instruments could be violated if locations along least cost road construction paths between major economic centers in China are correlated with economic county characteristics due to history and sorting. Furthermore, the instrument is likely to be mechanically correlated with distance to the nearest targeted metropolis. I therefore estimate regressions before and after including a set of additional county controls that could be correlated with the instrument while also affecting changes in economic outcomes between 1997-2006.

Counties closer to targeted city centers are mechanically more likely to lie on a least cost spanning tree path than counties situated farther away. Concerns about the IV strategy arise if distances to the major cities of China are correlated with economic county characteristics that also affect growth trajectories. I therefore include the log distance between counties and the nearest targeted metropolitan city center to address this concern.

Conditional on county distance to the targeted centers, location on least cost road construction paths between major economic centers in China could be correlated with political and economic county characteristics due to historical trade routes. To address such concerns, I include a set

[9] As discussed in the following section, I also include these geographical characteristics as county controls to address the concern that they might affect changes in economic outcomes directly.

[10] See Online Appendix for details and maps.

[11] I define these geographical areas on the basis of six geographic regions with official administrative recognition in China: East, North, North-East, North-West, South-Central, and South-West.

of observable controls for pre-existing county level political status and economic conditions. The two political controls are dummy variables indicating whether the county seat was a prefectural capital or has city rather than township status in 1990. The concern is that higher administrative status might be historically concentrated along least cost path routes between important economic centers.

Concerning pre-existing economic conditions, I use data from the 1990 Census at the county level which allows me to compute the share of agricultural employment in total county employment, the logarithm of county level urban registered population, as well as the share of above compulsory schooling attainment in total county population above 20 years of age in 1990.[12] These controls are aimed to address concerns that counties along least cost connections between major cities differ in terms of both their economic composition (shares of skilled labor and sectoral specialization), as well as in their mass of economic activity (urban populations).

The baseline identifying assumption is that county location along an all-China least cost spanning tree network affects changes in county level economic outcomes only through NTHS highway connections, conditional on province fixed effects, distance to the nearest targeted city region, administrative status and county-level economic conditions in 1990. The following section reports and discusses the baseline estimation results and a number of additional robustness checks to assess the validity of this assumption.

4 Estimation Results

This section reports estimation results and robustness checks of specification (1) for a number of different county level economic outcomes. Table 2 presents the first stage results for the least cost path and the Euclidean network instruments as well as their combined first stage results. I report first stage results for both binary NTHS connection indicators and for the log distance to the nearest NTHS segment. Both the least cost path and the straight line networks are strongly significant within province predictors of actual NTHS placements conditional on log distance to the nearest targeted node and the full set of pre-existing political and economic county characteristics. County controls are related to NTHS exposure mostly as expected. NTHS route connections are more likely for counties with lower distances to the targeted metropolitan city centers, larger pre-existing urban populations, and city status.[13]

Both instruments remain statistically significant when included simultaneously, confirming that the two spanning tree networks capture slightly different sources of the increased likelihood of route placements. While the least cost path instrument is a more precise predictor of placements on any given bilateral connection, the Euclidean tree instrument captures a higher proportion of the actually built network connections.

[12] Categories beyond the compulsory 9-year curriculum are senior middle school, secondary technical school, technical college, junior college and university.

[13] When individually included, higher pre-existing shares of agricultural employment decrease the likelihood of route placements, and higher shares of educated population increase it. Due to collinearity these associations are no longer significant when both controls are added simultaneously as reported in Table 2. Finally, the identifier for prefecture level capital status in 1990 enters in opposite sign than expected (decreased likelihood). This is due to the simultaneous inclusion of the city identifier, so that the coefficient is driven by approximately 1% of relatively remote prefecture level capitals that are not also classified as cities in 1990.

Table 3 presents OLS and IV results when regressing log changes of county level outcomes on the binary network treatment variable before and after including the full set of 1990 county controls.[14] The instrumental variable estimates of the NTHS connection effect are negative and statistically significant for industrial output growth, non-agricultural output growth, local government revenue growth, as well as total GDP growth.

Two important patterns emerge from Table 3. The first is that the IV point estimates are more negative than the OLS estimates, which are negative but not statistically significant in most cases. The second is that the inclusion of additional county controls for pre-existing political status and economic conditions leads to more negative point estimates of the NTHS connection effect. These findings are in line with the discussed concern that planners targeted places with higher expected returns to infrastructure investments and/or higher expected traffic demand, which is also apparent in the descriptive statistics of Table 1. These findings are also consistent with the results of Baum-Snow *et al.* (2012) who also report evidence suggesting that Chinese transport planners targeted economically prosperous regions.

The above results suggest that county location along least costly road construction paths between major cities in China is at least partly correlated with pre-existing county characteristics, such as the size of urban populations, the share of educated labor, and the degree of industrialization. As noted in the previous section, these correlations could be driven by settlements and sorting along historical trade routes. However, the finding that conditioning on pre-existing county characteristics leads to more negative connection effects on industrial output growth, total output growth, as well as local government revenue growth suggests that these characteristics are positively associated with economic growth, rather than negatively.

Nevertheless, the sensitivity of the IV point estimates to the inclusion of county controls in principle raises the concern that the estimated effects remain biased in either direction due to omitted unobserved differences that are correlated with the instrument. A related concern is that counties along an all-China spanning tree network had different pre-existing growth trends before the highway network came into effect. To address such concerns, I make use of the fact that the majority of the reporting county sample in 1997 and 2006 also reported local government revenues in the Provincial Statistical Yearbooks for the year 1990. If the instruments are valid conditional on the included county controls, then we should expect to find no significant relationship between NTHS treatments and local government revenue growth prior to the network, when estimated on the identical county sample for both periods.

Table 4 presents OLS and IV results for both instruments in both periods 1990-1997 and 1997-2006. For completeness, the table also reports results for the continuous NTHS exposure variable measured by the log county distance to the nearest NTHS route for both periods. The county sample is smaller than in the previous regressions firstly because not all reporting counties in 1997-2006 have non-missing entries for local government revenue in 1990, and secondly because these regressions exclude counties that were connected to NTHS routes built between 1992-1997 (10% of the NTHS).

The connection indicator enters negatively and statistically significantly only for the NTHS

[14]The Online Appendix reports results after replacing the binary network treatment with the log county distance to the nearest NTHS segment. The results discussed in this section are confirmed in opposite sign as expected.

period, and the log distance to the nearest NTHS segment enters positively and statistically significantly only for the NTHS period. Furthermore, these results appear to be not driven by differences in the size of the standard errors of the point estimates across the two different periods, but by changes in the point estimates themselves. These findings provide some additional reassurance on the validity of the instruments conditional on the included set of pre-existing political and economic county controls.[15]

In terms of magnitudes of the estimated effects, the IV estimates using both spanning tree instruments and the full set of county controls as reported in the final column of Table 3 suggest that NTHS connections on the way between targeted destinations have on average reduced GDP growth by about 18 percent over a nine year period between 1997-2006 compared to non-connected peripheral counties. Local government revenue growth has been reduced by approximately 23 percent. These adverse effects appear to be mainly driven by a decline in industrial output growth of approximately 26 percent over the nine year period.[16]

The results for agricultural GDP growth are close to zero and not statistically significant. This finding would point against a reallocation of factors of production from industry to agriculture. This result could be due to labeling both more and less industrialized activities as agricultural in county level economic accounts. Alternatively, the result could be driven by factor market rigidities and related to the frequent empirical finding that the inter-sectoral reallocation of resources following trade shocks fails to be confirmed in the data (Goldberg and Pavcnik, 2007; Topalova, 2010).

The final result from Table 3 is that NTHS connections appear to have no significant effect on county population growth. The point estimates are close to zero and statistically insignificant conditional on controls. This result appears to be consistent with Chinese migration controls under the hukou system (e.g. Au and Henderson, 2006), and suggests that the estimated effects on output growth are not driven by significant differences in population growth across counties.

In addition to the results reported in this section, the Online Appendix includes a series of additional robustness checks concerning the estimated average NTHS connection effects on production and government revenue growth. The additional results address potential concerns about i) controlling for direct effects of the land cover and elevation characteristics used in the construction of the least cost path instrument, ii) controlling for construction activity already underway in 1997, iii) excluding mountainous regions due to least cost path endogeneity concerns, iv) excluding regions along the Golmud-Lasa railway route completed over the same period, v) controlling for proximity to historical trade routes such as the Silk Road, vi) controlling for 1997 county differences in distance weighted market access, vii) controlling for county proximity to the coastline, and viii) testing for sample selection concerns due to boundary changes. In summary, the results discussed in this section appear to be robust in their magnitude and statistical significance.

The Online Appendix also provides a discussion and estimation results concerning the proportion and observable characterization of the complier group of counties that drive the estimated

[15]The system of government revenue collection underwent reforms in 1994 under the so called tax sharing system (Wong, 2000, Qiao *et al.*, 2008; Lin, 2009). As discussed in the Online Appendix, these reforms are unlikely to have affected local government revenue growth differently among county locations on a minimum spanning tree network.

[16]The magnitudes correspond to point estimates as shown in the regression tables after converting log points back to percentage changes and rounding.

local average treatment effects. Descriptive statistics and the pattern of coefficient estimates discussed above suggest that planners targeted economically prosperous counties on the way between targeted city regions. The additional estimations address the possibility that least cost spanning tree location might have affected actual highway placements only for a subset of remote and economically stagnant counties on the way between targeted nodes, so that the estimated local average effects might systematically differ from population average effects. The results presented in the Online Appendix suggest little support for this concern, suggesting that the predictive power of the instruments does not significantly vary across observable pre-existing county characteristics. The reasons behind this finding appear to be due to the nature of the spanning tree prediction errors compared to actual NTHS route placements, which I further explore and illustrate in the Online Appendix.

5 Channels at Work

The preceding section has presented empirical evidence suggesting that NTHS network connections have led to reduced GDP growth among peripheral counties on the way between targeted metropolitan city nodes relative to non-connected peripheral regions. These effects appear to be mainly driven by a reduction in industrial output growth. This section provides a discussion and additional estimation results to further investigate the channels at work. The main question that I address is whether the observed effects are driven by a mechanism based on inter-regional trade cost reductions between connected peripheral and metropolitan regions, as opposed to a mechanism based on the local effects of the NTHS on decentralization from connected to neighboring non-connected peripheral counties along the network routes.

On one side, the observed results could be driven by inter-regional trade integration. In particular, the results could be accounted for by core-periphery effects of trade integration between *ex ante* asymmetric markets as found in the literature on increasing returns trade theory and economic geography. In the presence of increasing returns to scale in industrial production, Dixit-Stiglitz monopolistic competition, and iceberg trade costs, Krugman (1980) and Helpman and Krugman (1985) have provided a microfoundation for the proposition that market size is a determinant of industrialization, and that falling trade costs between large and small markets can reinforce the concentration of production in the larger market.[17]

This result can either be driven solely by the so called home market effect as in Krugman (1980) and Helpman and Krugman (1985), or by an interplay of forces including additional self-reinforcing agglomeration forces that arise in the context of labor mobility as introduced by Krugman (1991). In this literature, the common foundation of the core-periphery effect of trade integration is that while falling trade costs reduce the strength of both agglomeration forces (such as better access to consumers or to intermediate inputs in larger markets) and dispersion forces (such as less product market competition or lower factor prices in smaller markets), they tend to attenuate the agglomeration forces at a lower rate compared to the dispersion forces.

A trade based mechanism could also be based on neoclassical constant returns to scale trade

[17]See Fujita *et al.* (1999), Baldwin *et al.* (2003), and Combes *et al.* (2008) for comprehensive reviews of this literature.

theory. In the presence of locational fundamentals that give rise to a comparative disadvantage in industrial production among peripheral regions relative to metropolitan centers, falling trade costs would lead to a decline in industrial production among connected peripheral regions. In turn, total production could be adversely affected in this setting in the presence of either frictions in the reallocation of factors of production across sectors (Goldberg and Pavcnik, 2007; Topalova, 2010), or due to outflows of factors used in industrial production (e.g. capital) (Redding, 2012).

On the other hand, the observed results could be driven by the local effects of the NTHS within peripheral regions along the network routes, rather than by inter-regional trade cost reductions between peripheral and metropolitan regions. In particular, this channel could be motivated by recent findings in Baum Snow *et al.* (2012) in the urban economics literature. Using rich historical geographical information about China's urbanization process, they document that a process of urbanization and industrial decentralization has taken place over recent decades among the central city districts of Chinese prefecture level capitals. That is, while population has grown at higher rates in central cities compared to their urban peripheries, the opposite has been the case for industrial production growth. In this context, they present evidence that road and railway infrastructure can significantly affect this process of urbanization and decentralization within metropolitan areas.

These findings could be relevant in two central ways for the interpretation of the presented reduced form effects of NTHS connections among peripheral counties in China. First, NTHS routes could have a causal effect on the urbanization and industrial decentralization from connected peripheral counties to non-connected peripheral counties in their neighborhood. Alternatively, it could be the case that the estimation sample of non-targeted peripheral counties include a significant fraction of second tier cities which undergo a process of urbanization and decentralization, while least cost path location along the all-China spanning tree instrument could be correlated with the location of such city centers. In other words, the IV assumptions could be violated because the instrument picks up decentralizing city centers relative to surrounding peripheral counties. In both scenarios, the negative reduced form effects reported in the previous section would be driven by a decentralization of industrial production from connected peripheral regions to neighboring areas, as opposed to the effects of trade integration between large metropolitan city regions and connected peripheral regions under the trade based channel.

While in principle each of the above mechanisms could give rise to the estimated reduced form effects on peripheral county outcomes, the channels would also imply different testable implications on a series of additional observable outcomes. The remainder of this section aims to provide further empirical evidence on these additional moments.

5.1 Additional Estimation Results

The first three additional estimations are aimed to differentiate between a trade based channel and a mechanism based on local decentralization. In particular, a mechanism based on urbanization and decentralization would imply that peripheral NTHS connections are correlated with increases in urbanization among connected peripheral counties relative to non-connected ones. It would also imply that counties in close proximity to the connected peripheral markets are affected more positively by NTHS connections in terms of industrial production and GDP relative to peripheral counties farther away from the direct neighborhood of the connected counties. Neither of these predictions would hold if inter-regional trade integration between metropolitan and peripheral regions was driving the results. Furthermore, while a trade based channel would imply that the adverse peripheral connection effect is decreasing in county distance to the nearest NTHS route, the local decentralization channel predicts a non-monotonicity in the relationship between the network connection effect and county distance to the NTHS.

Table 5 presents the estimation results on these three testable implications. Columns 1 and 2 report the estimated NTHS treatment effect on urban population growth as well as changes in urbanization among connected peripheral counties after using both least cost path and Euclidean spanning tree instruments to instrument for route placements. The finding that the IV point estimates are close to zero and statistically insignificant in both cases presents initial evidence that a process of urbanization among connected peripheral counties is unlikely to underlie the reduced form effects presented in the previous section.

The remaining Columns 3-11 of Table 5 further investigate the spatial pattern of the estimated NTHS connection effects among non-targeted peripheral counties. In particular, Columns 4, 7, and 10 report the IV estimates of the NTHS treatment effect on growth in industrial production, GDP and government revenues respectively after including an additional indicator variable which takes the value of one for peripheral counties that are next to connected peripheral counties and zero otherwise.[18] This specification implies a partition of the peripheral county sample into connected counties, neighboring counties, and counties farther away from connected counties which represent the new reference category in the estimation.

The finding that the baseline NTHS effect reported in Columns 3, 6, and 9 is virtually unchanged by the inclusion of the additional peripheral neighbor dummy does not support the prediction that surrounding peripheral areas are more positively affected by NTHS connections compared to counties farther away. Instead, the finding suggests that the estimated adverse effect of network connections is concentrated within connected county regions.[19]

The final columns of Table 5 and Figure 4 confirm this finding of a relatively steeply declining adverse connection effect over distance to the nearest NTHS route with respect to industrial output growth, GDP growth, as well as local government revenue growth. Columns 5, 8 and 11 report a statistically significant positive effect of log distance to the nearest NTHS route placement for all three dependent variables.[20] Figure 4 then plots the best fitting polynomial functional forms of the flexibly estimated relationships between county distance to the nearest NTHS segment and industrial output growth, GDP growth, or government revenue growth.[21] The graphs suggest a relatively steeply declining monotonic relationship of the adverse connection effect over distance to the NTHS, and once again there is no evidence of a non-monotonicity in this relationship around the median distance of peripheral counties in the direct neighborhood of connected counties. To

[18] To instrument for the additional neighbor dummy, I use both an indicator for adjacency to a connected county on the LCP spanning tree and an indicator for adjacency to a connected county on the Euclidean spanning tree.

[19] As pointed out by a referee, these results are also informative with respect to so called agglomeration shadows (Fujita and Krugman, 1995; Chandra and Thomson, 2000). In particular, I do not find evidence suggesting that connected peripheral counties attract industrial activity from neighboring non-connected counties.

[20] To instrument for distance to the nearest NTHS segment, I use both distance to the nearest LCP spanning tree and the nearest Euclidean spanning tree as instruments.

[21] The plotted relationships correspond to the best fitting specification according to the Akaike Information Criterion (AIC). Distance polynomials are instrumented with distances to LCP and Euclidean spanning trees.

summarize, the findings reported in Table 5 and Figure 4 provide evidence that appears consistent with a trade based channel, while they provide little support for a mechanism based on local decentralization among non-targeted peripheral regions in the context of the Chinese NTHS.

In relation to existing findings in the urban economics literature by Baum-Snow (2007) and Baum-Snow *et al.* (2012), these findings suggest that large scale inter-regional transport networks can lead to a different set of economic effects than those found when analyzing the local effects of transport infrastructure on urban form within metropolitan areas. This difference in policy questions also translates into different empirical settings. One such difference is the spatial scale of the analysis. The current empirical strategy is aimed at estimating the effects of NTHS connections among non-targeted peripheral counties outside the commuting zones of the targeted metropolitan city regions. On the other hand, studies of the local within-metropolitan effects of road additions set focus on comparisons between central city districts and surrounding urban areas within metropolitan regions.[22] In this context, the presented findings also relate to a longstanding argument in the trade literature that stresses the conceptual differences when evaluating the effects of large scale inter-regional transport infrastructure as opposed to the local within-region effects of transport policies (e.g. Ottaviano, 2008).

The final estimation results are aimed at testing a set of additional observable implications that arise from a simple core-periphery model of trade integration between asymmetric regions. These results concern the observed heterogeneity of the NTHS connection effect across peripheral counties with different pre-existing characteristics. The first source of heterogeneity is that the adverse connection effect should be less pronounced among peripheral counties with larger pre-existing market sizes. The second source of heterogeneity is that, holding constant initial differences in market sizes, the adverse connection effect should be less pronounced among peripheral counties whose initial level of trade costs *vis-à-vis* the metropolitan core regions is higher. The Online Appendix provides a formal exposition of these testable implications within a simple increasing returns to scale trade model based on Helpman and Krugman (1985).[23] The key intuition behind both interaction effects is that falling trade costs attenuate the locational advantage of producing in a remote location at a faster rate than the agglomeration advantage of producing in the core region, so that for a more pronounced initial core-periphery size gradient, or for lower initial trade costs, a given reduction in trade costs induces more industrial concentration in the core regions.[24]

To test the heterogeneity of the network's effect among peripheral counties, I estimate specification (1) after including interaction terms between the NTHS connection indicator and a number of theoretically motivated observable pre-existing county characteristics. Alternatively, I replace the binary NTHS connection indicator in these interactions by the continuous treatment variable of log distance to the nearest NTHS route as in the previous estimations reported in Table 5

[22] Another important difference is the county population under study. In the case of China, only 30 (28) of the 295 prefecture level capitals within Han Chinese provinces that make up the estimation sample in Baum-Snow *et al.* (2012) are located on the least cost path spanning tree network (Euclidean spanning tree network), and prefecture capitals account for less than 10% of the peripheral county estimation sample.

[23] The model is a multi-region version of Helpman and Krugman (1985) that also allows for mobile capital across Chinese counties. In line with the empirical setting, the predictions are derived for differential changes in outcomes between connected and non-connected peripheral regions in the light of falling trade costs *vis-à-vis* metropolitan centers.

[24] These results are related to the concept of home market magnification in the trade literature (Baldwin *et al.*, 2003). See Online Appendix for discussion.

and the Online Appendix. In particular, county log distance to the nearest targeted metropolitan node and an indicator for counties with above mean 1990 total employment sizes are aimed to capture the initial degree of trade costs between targeted core cities and peripheral counties and the pre-existing degree of core-periphery size differentials respectively.

Table 6 reports second stage IV estimation results for industrial output growth as well as total GDP growth as dependent variables after instrumenting for NTHS connection and its interaction terms, or for log NTHS distance and its interaction terms, with the least cost path network instrument.[25] The first column for each dependent variable reproduces the average treatment estimate conditional on the full set of county controls as reported in Table 3. The second columns then introduce the additional two interaction terms motivated by the core-periphery channel. In the binary network connection specifications both interaction terms enter positively and statistically significantly for both industrial output growth and GDP growth. When using the log distance to the nearest NTHS segment as continuous treatment variable instead, the interaction terms enter negatively and statistically significantly for both dependent variables.[26]

The reported results provide empirical evidence of significant heterogeneity in the NTHS connection effects among peripheral counties. In particular, while the majority of both large and small peripheral counties are estimated to have experienced negative GDP growth effects, a subset of large peripheral counties that are also subject to large initial trade costs with respect to the targeted metropolitan city centers are estimated to have experienced marginally positive growth effects due to NTHS connections.[27] These results contribute to a richer understanding of the estimated effects in Section 4, and they appear to be consistent with the predictions of a simple core-periphery trade model based on the home market effect in Helpman and Krugman (1985). On the other hand, the results of Table 6 could also be rationalized within a model of inter-regional trade integration based on constant returns to scale and comparative advantages. In this respect, it would be beyond the scope of the current empirical setting to test for one particular trade channel and rule out alternative microfoundations.[28]

To summarize, the estimation results reported in Tables 5 and 6 and Figure 4 provide additional empirical evidence of the NTHS connection effects on urbanization, decentralization of production, as well as the heterogeneity of the effects with respect to pre-existing county characteristics. In conclusion, the presented findings provide additional qualitative evidence suggesting that falling trade costs between *ex ante* asymmetric peripheral and metropolitan regions are driving the observed effects.

[25] In the cross-derivative regressions the least cost path spanning tree instrument performs better in terms of first stage predictive power than the Euclidean instrument or both instruments combined. The Online Appendix reports these specifications using both instruments instead.

[26] A potential concern with these results is that the first stage F-statistics significantly drop after instrumenting for NTHS treatments as well as their interactions. The Online Appendix reports an additional set of results that compare two stage least squares (2SLS) estimates using both instruments to estimations by limited information maximum likelihood (LIML). The fact that the reported LIML point estimates are slightly higher than the 2SLS estimates reported in the Online Appendix indicates that weak instrument bias is unlikely to be a concern. See for example Angrist and Pischke (2008, Section 4.6).

[27] The median and mean distances to the nearest targeted city node are 168 and 203 km respectively.

[28] In particular, the first stage F-statistics already significantly drop in columns 2 of Table 6. The earlier working paper version of this paper also reports results after including additional interaction terms that are aimed to further distinguish between different trade channels.

6 Conclusion

Large scale transport investments connect both large metropolitan centers of production as well as small peripheral market places. This is especially the case in developing countries where spatial disparities are particularly pronounced. A common policy view is to combine national efficiency objectives with regional equity objectives of transport policies under the presumption that falling inter-regional trade costs promote both national gains from trade as well as the diffusion of economic activity to peripheral regions. This paper exploits China's NTHS as a source of plausibly exogenous variation in trade cost shocks across a large number of *ex ante* asymmetric regions to provide empirical evidence on this question.

The presented findings provide empirical evidence that large scale inter-regional transport infrastructure can lead to a reduction in industrial and total output growth among connected peripheral regions relative to non-connected ones, rather than diffusing production from metropolitan regions to the periphery. Additional estimation results present further qualitative evidence in support of a mechanism based on falling trade costs between peripheral and metropolitan regions. These findings relate to existing theoretically motivated policy discussions in the trade literature (Fujita *et al.*, 1999; Baldwin *et al.*, 2003; Combes *et al.*, 2008), and serve to emphasize the importance of potentially unintended general equilibrium consequences when evaluating and planning large scale transport infrastructure policies.

Finally, it is important to emphasize that the presented findings are perfectly consistent with potentially large aggregate efficiency improvements. Each of the discussed channels underlying the estimated reduced form effects would also imply aggregate gains from trade. In this context, it is also important to point out some of the main limitations of the presented analysis. In particular, while the NTHS policy design in combination with the proposed instrumental variable strategy have the advantage to exploit plausibly exogenous variation in network connections among non-targeted peripheral counties, the current empirical setting would not be suited to evaluate the network's economic impact at the national level.

References

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent.* Cambridge, Mass.: Harvard Univ. Pr.

Angrist, J., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Asian Development Bank. (2007). A retrospective analysis of the road sector 1997-2005. *Asian Development Bank Operations Evaluation Department Report.*

Asian Infrastructure Monthly. (1995). China paves the way for expressway network. *Asian Infrastructure Monthly, June.*

Au, C., & Henderson, J. (2006). How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics*, *80*(2), 350–388.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., & Robert-Nicoud, F. (2003). *Economic geography and public policy.* Princeton University Press.

Banerjee, A., Duflo, E., & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China. *Yale Department of Economics mimeo.*

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quaterly Journal of Economics*, *122*(2), 775.

Baum-Snow, N., Brandt, L., Henderson, J., Turner, M., & Zhang, Q. (2012). Roads, railroads and decentralization of Chinese cities. *mimeo, Brown University*.

Baum-Snow, N., & Turner, M. (2012). Transportation and the decentralization of Chinese cities. *mimeo, Brown University*.

Behrens, K., Lamorgese, A., Ottaviano, G., & Tabuchi, T. (2009). Beyond the home market effect: Market size and specialization in a multi-country world. *Journal of International Economics*, *79*(2), 259–265.

Brülhart, M., & Trionfetti, F. (2009). A test of trade theories when expenditure is home biased. *European Economic Review*, *53*(7), 830–845.

Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, *30*(4), 457–490.

CIBC. (2009). Capitalizing on the upcoming infrastructure stimulus. *CIBC World Markets Occasional Report No. 66*.

Combes, P., Mayer, T., & Thisse, J. (2008). *Economic geography: The integration of regions and nations*. Princeton University Press.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, *92*(1), 1–45.

Davis, D. (1998). The home market, trade, and industrial structure. *American Economic Review*, *88*(5), 1264–76.

Davis, D., & Weinstein, D. (1996). Does economic geography matter for international specialization? *National Bureau of Economic Research Working Paper*.

Davis, D., & Weinstein, D. (1999). Economic geography and regional production structure: An empirical investigation. *European economic review*, *43*(2), 379–407.

Davis, D., & Weinstein, D. (2003). Market access, economic geography and comparative advantage: An empirical test. *Journal of International Economics*, *59*(1), 1–23.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, *1*(1), 269–271.

Donaldson, D. (2013). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review, forthcoming*.

Duranton, G., & Turner, M. (2012). Urban growth and transport. *Review of Economic Studies*, *79*, 1407–1440.

Fujita, M., & Krugman, P. (1995). When is the economy monocentric?: von Thunen and Chamberlin unified. *Regional Science and Urban Economics*, *25*(4), 505–528.

Fujita, M., Krugman, P., & Venables. (1999). *The spatial economy: Cities, regions and international trade*. Wiley.

Garske, T., Yu, H., Peng, Z., Ye, M., Zhou, H., Cheng, X., & Ferguson, N. (2011). Travel patterns in China. *PloS one*, *6*(2).

Goldberg, P., & Pavcnik, N. (2007). Distributional effects of globalization in developing countries. *Journal of Economic Literature*, *45*, 39–82.

Hanson, G., & Xiang, C. (2004). The home-market effect and bilateral trade patterns. *American Economic Review*, 1108–1129.

Head, K., & Ries, J. (2001). Increasing returns versus national product differentiation as an explanation for the pattern of US-Canada trade. *American Economic Review*, 858–876.

Helpman, E., & Krugman, P. (1985). Market structure and international trade. *Cambridge (Mass.)*.

Jha, M., McCall, C., & Schonfeld, P. (2001). Using GIS, genetic algorithms, and visualization in highway development. *Computer-Aided Civil and Infrastructure Engineering*, *16*(6), 399–414.

Jong, J., & Schonfeld, P. (2003). An evolutionary model for simultaneously optimizing three-dimensional highway alignments. *Transportation Research Part B: Methodological*, *37*(2), 107–128.

Kanbur, R., & Venables, A. J. (2005). *Spatial inequality and development.* Oxford University Press.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review*, *70*(5), 950–959.

Kruskal, J. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, *7*(1), 48–50.

Li, S., & Shum, Y. (2001). Impacts of the national trunk highway system on accessibility in China. *Journal of Transport Geography*, *9*(1), 39–48.

Lin, S. (2009). The rise and fall of China's government revenue.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, *90*(4), 683–701.

Muth, R. (1969). *Cities and housing: The spatial patterns of urban residential land use.* University of Chicago, Chicago.

Ottaviano, G. (2008). Infrastructure and economic geography: An overview of theory and evidence. *European Investment Bank Papers*.

Puga, D. (2002). European regional policies in light of recent location theories. *Journal of Economic Geography*, *2*(4), 373.

Qiao, M.-V. J., B., & Xu, Y. (2008). The tradeoff between growth and equity in decentralization policy: China's experience. *Journal of Development Economics*, *86*(1), 112–128.

Redding, S. J. (2012). Goods trade, factor mobility and welfare. *NBER Working Paper*(w18008).

Redding, S. J., & Sturm, D. M. (2008, December). The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, *98*(5), 1766-97.

South China Morning Post. (1994). Construction on the road to success. *South China Morning Post*, *September 08*.

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics*, *2*(4), 1–41.

Wong, C. (2000). Central-local relations revisited the 1994 tax-sharing reform and public expenditure management in China. *China Perspectives*, 52–63.

World Bank. (2007a). China study tour by New Delhi transport office team. *World Bank Internal Report*.

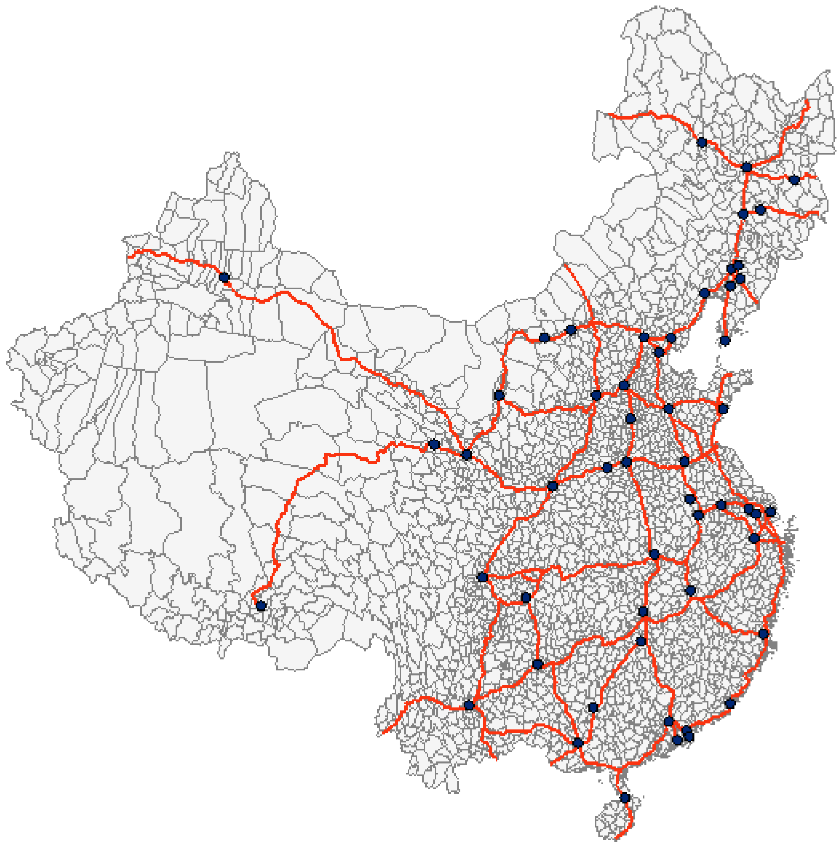
World Bank. (2007b). Domestic trade impacts of the expansion of the national expressway network in China. *World Bank EASTR Working Paper No. 14*.

World Bank. (2008). Safe, clean, and affordable: Transport for development. *World Bank Transport Business Strategy 2008-2012*.

Appendix

Figures

Figure 1: **China's National Trunk Highway System**

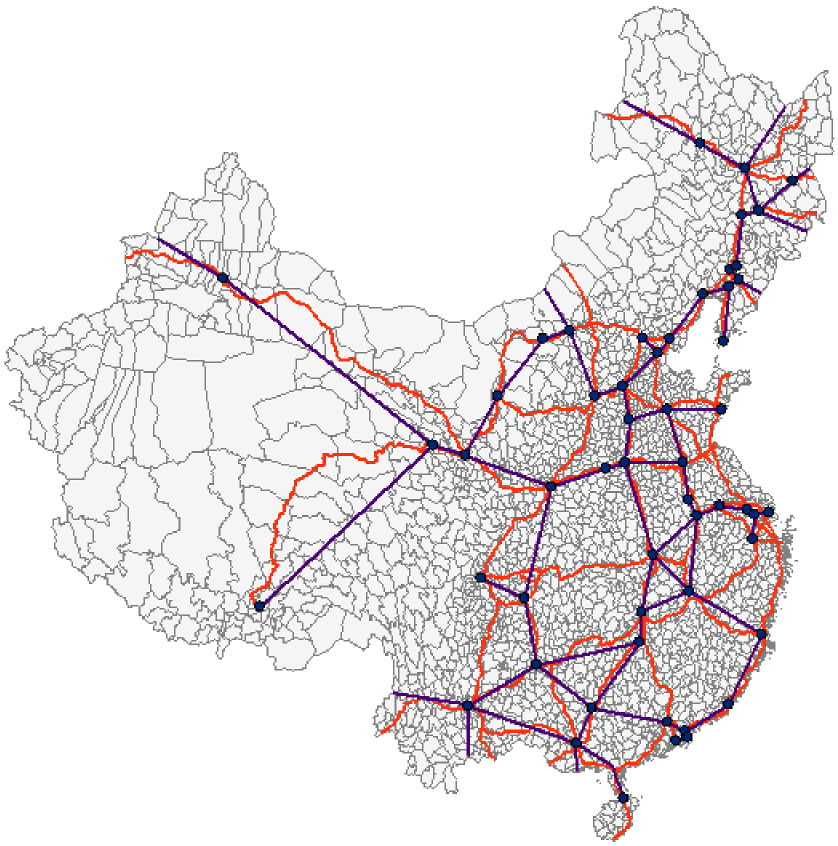


The figure shows Chinese county boundaries in 1999 in combination with the targeted city nodes and the completed expressway routes of the National Trunk Highway System (NTHS) in the year 2007.

Figure 2: **Least Cost Path Spanning Tree Network**

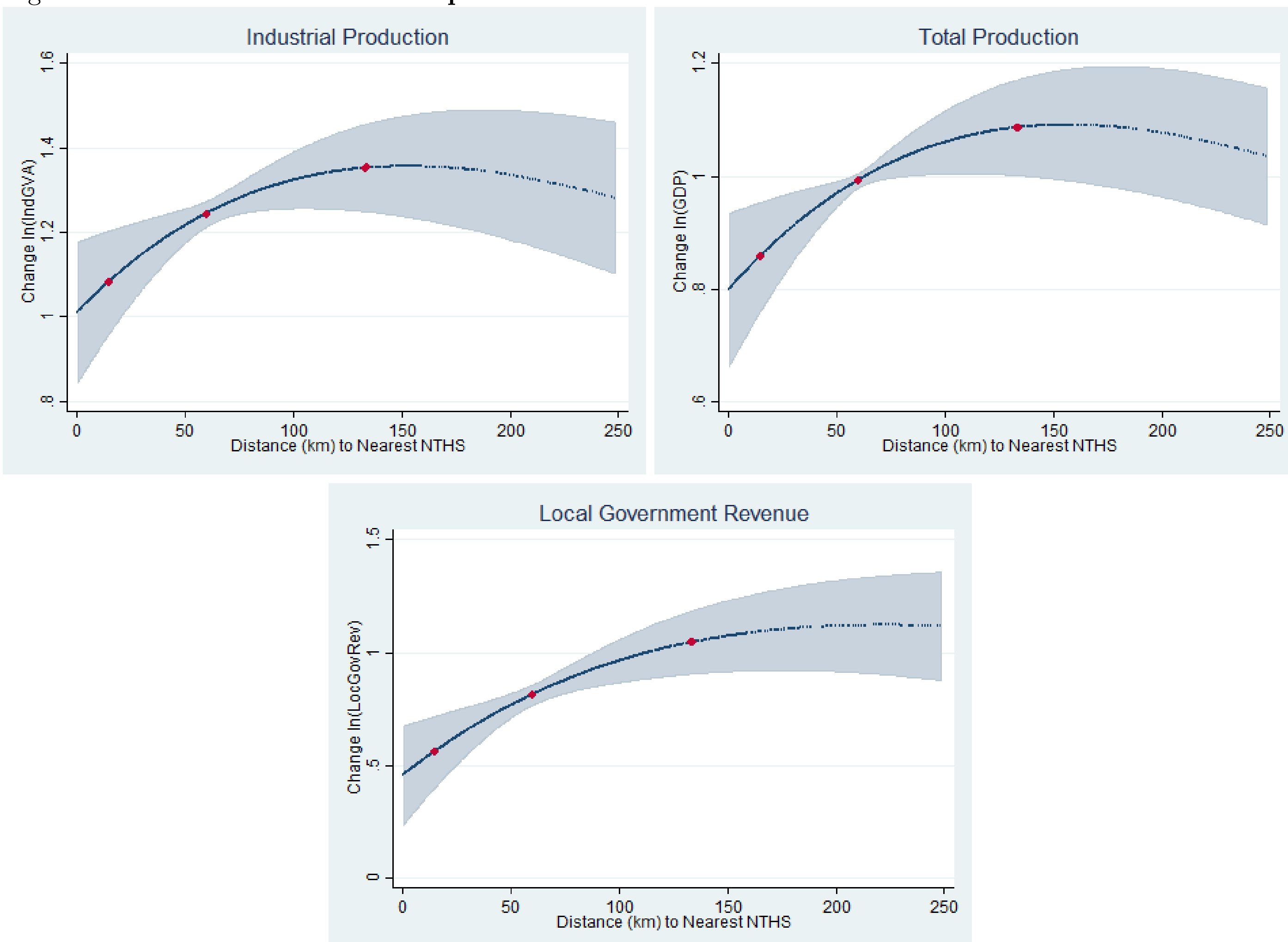
The network in red color depicts the completed NTHS network in 2007. The network in black color depicts the least cost path spanning tree network. The black routes are the result of a combination of least cost path and minimum spanning tree algorithms. In the first step Dijkstra's (1959) optimal route 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.

Figure 3: **Euclidean Spanning Tree Network**



The network in red color depicts the completed NTHS network in 2007. The network in darker color 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-Center-South and East-Center-West divisions of China. These regional repetitions add 9 routes to the original minimum spanning tree.

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 neighboring a connected county (center), and the remaining peripheral counties farther away (right). The shaded areas indicate 90% confidence intervals. Standard errors are clustered at the province level.

Tables

Table 1: **Descriptive Statistics for 1997**

	Targeted City Centers	Connected Periphery	Non-connected Periphery	National Share of Targeted City Centers
Population (10,000)	233.24	56.96	38.48	0.17
Urban population (10,000)	179.69	10.77	5.83	0.5
GDP (100 Million Yuan)	517.86	32.58	15.09	0.5
GDP per capita (Yuan)	21435.06	5142.16	3637.09	-
Local government revenue (100 Million Yuan)	38.23	1.23	0.57	0.67
Industrial gross value added (100 Million Yuan)	194.61	14.93	5.58	0.48
Non-Agricultural gross value added (100 Million Yuan)	505.75	24.42	9.74	0.59
Agricultural output share	0.04	0.34	0.42	-
Land area (km^2)	1543.09	3057.47	4513.4	0.015
Number of counties	54	424	943	54

The first three columns present mean 1997 levels, and the fourth column presents national shares by county groups. Targeted city centers refer to the central city county units (shixiaqu) of targeted metropolitan regions. Peripheral counties are counties outside a 50 km commuting buffer around the targeted city centers.

Table 2: **First Stage Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Connect	Connect	Connect	lnDistHwy	lnDistHwy	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

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: **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)	Connect	-0.0529	-0.0356	-0.284**	-0.304**	-0.246*	-0.287*	-0.272***	-0.297***	-0.297**
1997-2006		(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)	Connect	-0.0411	-0.0266	-0.243**	-0.252**	-0.270**	-0.296**	-0.251***	-0.268***	-0.268***
1997-2006		(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)	Connect	-0.0497*	-0.0914***	-0.0542	-0.223*	-0.175	-0.315**	-0.0926	-0.257***	-0.257***
1997-2006		(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)	Connect	-0.00204	-0.0144	-0.106	-0.177*	-0.178	-0.254**	-0.127	-0.203**	-0.203**
1997-2006		(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)	Connect	-0.00344	-0.00790	0.000194	-0.0252	-0.0305	-0.0597	-0.00865	-0.0371	-0.0371
1997-2006		(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)	Connect	0.00488	-0.00217	0.0395**	0.0264	0.0183	0.0104	0.0333*	0.0207	0.0207
1997-2006		(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							

Each point estimate stems from a separate regression. All regressions include province fixed effects. LCP IV stands for the least cost path spanning tree instrument. Euclid IV stands for the straight line spanning tree instrument. No Controls columns refer to regressions on NTHS treatment and log county distance to the nearest targeted city node. With Controls indicates a full set of 1990 county controls (city status and prefecture capital dummies, log urban population, share of agricultural employment, and share of above compulsory school attainment in 20+ population). The dependent variables in order as reported are county level industry gross value added, manufacturing plus services gross value added, local government revenue, total GDP, agricultural gross value added, and population. Column (9) reports Conley (1999) standard errors, estimated by GMM to adjust for spatial dependence without imposing parametric assumptions. I allow for spatial dependence up to the distance given by the diameter of the average province area. The remaining standard errors are clustered at the province level and stated in parentheses below point estimates. ***1%, **5%, and *10% significance levels.

Table 4: **Falsification Test Before and After the Network Was Built**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	OLS	OLS	LCP IV	LCP IV	Euclid IV	Euclid IV	Both IVs	Both IVs
Change ln(LocGovRev)	1990-97	1997-06	1990-97	1997-06	1990-97	1997-06	1990-97	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

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.

Table 5: **Are NTHS Connections Associated to Urbanization and Industrial Decentralization among Peripheral Counties?**

Dependent Variable:	Change ln(UrbPop) 1997-06	Change ln(Urb/Pop) 1997-06	Change ln(IndGVA) 1997-06			Change ln(GDP) 1997-06			Change ln(GovRevenue) 1997-06		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Connect	0.0350	0.0137	-0.297***	-0.262**		-0.203**	-0.193**		-0.257***	-0.253***	
	(0.0953)	(0.0925)	(0.108)	(0.113)		(0.0886)	(0.0919)		(0.0996)	(0.0961)	
Neighbor				0.153			0.0907			0.0535	
				(0.214)			(0.132)			(0.195)	
lnDistHwy					0.113*			0.0845*			0.177***
					(0.0615)			(0.0480)			(0.0667)
First Stage F-Stat	13.374	13.374	18.886	5.016	13.852	17.425	4.84	12.989	19.055	5.383	13.879
Obs	1,072	1,072	1,280	1,280	1,280	1,272	1,272	1,272	1,285	1,285	1,285

All regressions include province fixed effects and a full set of county controls. Reported results are 2nd stage estimates using the least cost path and the Euclidean spanning tree networks to instrument for NTHS connections, neighboring peripheral counties, or distance to the nearest NTHS segment. Columns 1 and 2 report connection effects on peripheral county changes in log urban population and urbanization respectively. Neighbor indicates peripheral counties neighboring a connected peripheral county. Standard errors are clustered at the province level and stated in parentheses below point estimates. ***1%, **5%, and *10% significance levels.

Table 6: **Testing the Heterogeneity of Peripheral Connection Effects**

Dependent variable:	Change ln(IndGVA) 1997-2006		Change ln(GDP) 1997-2006	
	(1)	(2)	(1)	(2)
Panel A: Binary				
Connect	-0.304**	-4.281***	-0.177*	-3.571***
	(0.145)	(1.569)	(0.0942)	(1.011)
Connect*ln(DistNode)		0.748***		0.636***
		(0.270)		(0.172)
Connect*Emp90Dum		0.450*		0.404**
		(0.255)		(0.196)
Obs	1280	1280	1272	1272
First stage F-Stat	29.966	3.462	27.972	4.724
Panel B: log Distance				
lnDistHwy	0.0954	1.465***	0.0639	1.105***
	(0.0674)	(0.455)	(0.0434)	(0.318)
lnDistHwy*ln(DistNode)		-0.236***		-0.181***
		(0.0748)		(0.0494)
lnDistHwy*Emp90Dum		-0.266***		-0.192***
		(0.0823)		(0.0693)
Obs	1280	1280	1272	1272
First stage F-Stat	22.367	4.649	21.698	4.842

All regressions include province fixed effects and a full set of county controls. Reported results are 2nd stage estimates using the LCP spanning tree to instrument for NTHS connections as well as their reported interaction terms. lnDistNode is log county distance to the nearest targeted city node. Emp90Dum is a dummy for counties with above mean levels of county employment in 1990. Standard errors are clustered at the province level and stated in parentheses below point estimates. ***1%, **5%, and *10% significance levels.